



THE MACMILLAN COMPANY
NEW YORK • BOSTON • CHICAGO • DALLAS
ATLANTA • SAN FRANCISCO

MACMILLAN AND CO., LIMITED
LONDON • BOMBAY • CALCUTTA • MADRAS
MELBOURNE

THE MACMILLAN COMPANY
OF CANADA, LIMITED
TORONTO

ELEMENTARY ECONOMICS

by

Fred Rogers Fairchild

KNOX PROFESSOR OF ECONOMICS

Edgar Stevenson Furniss

PELATIAH PERIT PROFESSOR OF SOCIAL SCIENCE

Norman Sydney Buck

PROFESSOR OF ECONOMICS

YALE UNIVERSITY

VOLUME I

Fourth Edition

THE MACMILLAN COMPANY

NEW YORK

1939

FOURTH EDITION COPYRIGHTED, 1939,
By THE MACMILLAN COMPANY.

All rights reserved — no part of this book may
be reproduced in any form without permission
in writing from the publisher, except by a re-
viewer who wishes to quote brief passages in
connection with a review written for inclusion in
magazine or newspaper.

Printed in the United States of America.
Set up and electrotyped Published May, 1939

Copyright, 1924 and 1925,
By Fred Rogers Fairchild
First, second, and third editions copyrighted and published 1926, 1930, and 1936,
By The Macmillan Company

PREFACE

This book represents the result of many years' experience in teaching economics to beginners at Yale University. That experience has convinced us of the importance of the textbook and has led us to certain definite conclusions as to the type of text best calculated to serve the purpose of helping college students to a knowledge of the fundamental facts and principles of the economic world in which they live. That is the sole purpose of this book. It is not written for the edification of the mature student of economics. It is strictly a book for beginners and the teachers of beginners.

Adherence to this purpose has indicated certain rules for our guidance and imposed certain restrictions upon us. As a rule the treatment has been confined to topics upon which the science of economics may fairly be said to have reached definite conclusions. Occasionally when this rule could not be followed we have stated, somewhat dogmatically perhaps, what we consider sound conclusions as to subjects upon which there may still be real difference of opinion among economists of authority. In a few cases we have undertaken an impartial summary of considerations upon the two sides of some unsettled current issue, but as a rule we do not believe that the beginning student profits greatly from that type of discussion which merely presents arguments on both sides of a controversial question and leaves him to draw his own conclusions. Such matters belong in the more advanced courses. There is plenty of material to occupy fully the elementary course in economics without penetrating far into those frontier regions where economic science has not yet reached fairly definite conclusions.

This limitation of the field makes easier the strictly scientific treatment to which we have sought to adhere. Our purpose is to present the facts and the principles of economics — to disclose the truth. With the ethical or moral problems which arise upon this economic foundation we are not concerned, except only as prevailing judgments upon such matters may themselves become the

subject matter of scientific economic inquiry. Our purpose is to aid the student to discover how things are, rather than how they ought to be. We have no case to establish, and we offer no propaganda.

Addressed as it is to beginning students, this is essentially an elementary book. We have sought to spare the student the confusion that comes from taking for granted knowledge which he cannot fairly be assumed to possess, and the treatment is correspondingly elementary. We have sought, by giving care to such things as definition of technical terms, sequence of topics, and simplicity of style, to enable the beginning student to follow the discussion without undue difficulty. It is not to be denied that certain parts of the subject of economics are difficult. We have made no attempt to escape such difficulties, either by avoiding the hard topic or glossing it over with a superficial discussion. Rather have we relied upon simple and elementary treatment to remove unnecessary causes of confusion, in the belief that so presented these topics will not prove beyond the capacity of the student.

The plan of this book has dictated the continuous combination of theoretical analysis with historical narrative and discussion of practical problems. It is our conviction that this is a sound pedagogical rule for an elementary treatise. Theory must be illustrated and justified by showing its relation to practical affairs; facts must be made significant and interesting by showing their conformity to general principles. "Theory" and "practice" should, we believe, go always hand in hand.

It is today quite generally recognized that the average college student does not possess that knowledge of the economic environment which is to be desired as a foundation for the study of economic principles and problems. A section (Part I) devoted to the study of the modern economic organization and its development, stressing the coöperative nature of modern society and the importance of price as a motivating force, is therefore made an integral part of this book.

The exercises which will be found at various points do not conform to any rigid rule. They have been introduced only at those points where it was felt that they would contribute substantial aid to the student in his effort to grasp the principles expounded

in the preceding text or would give useful training in economic reasoning. These problems are obviously illustrative rather than comprehensive, and the teacher can readily modify or expand them or ask the student himself to construct similar ones.

At the end of each of the major subdivisions of the book will be found a classified list of books which are suggested for the student's further reading. These lists are, of course, not comprehensive, and they purposely do not embrace works of a highly advanced or specialized or technical nature. They are intended to present for the choice of the student certain books which he will find interesting and profitable reading in connection with his first study of economics.

As compared with the majority of textbooks in economics, this book will appear quite long. It has been written with the idea of comprising the bulk at least of the reading to be assigned as the required work in a college course involving three exercises a week for a full year. We are of the belief that the present-day importance of economics justifies the devotion of that much time to the general elementary course.

These in brief are the more important principles which have guided us in harmony with our aim to give to those students who may not pursue the study of economics further a broad knowledge and understanding of the economic world of today and to others a firm foundation for their further study of economics and allied subjects. It is in such matters as are here noted that our contribution, if any, is to be found.

While we have had in mind predominantly the needs of college students, we are nevertheless bold enough to believe that the requirements thus imposed upon us will have made this book useful also to those more mature readers who, for any reason whatever, may have need of a simple elementary exposition of the fundamental facts and principles of economics in the light of present-day knowledge.

This book appeared first in 1926. It has undergone two thorough revisions, in 1930 and 1936. In these revisions we took advantage of the suggestions of many teachers who had used the book as well as of our own added experience in its use. In the 1930 revision the

treatment of price was strengthened by the addition of numerous examples from practical business, making this part less theoretical and mechanical and reducing the number of graphs required. Though the statement of essential principles remained unchanged, these principles were brought into closer relation to practical business affairs, and the discussion became somewhat less abstract than before. A section on mutually related prices was added. The 1936 edition brought in a more rigorous analysis of monopoly price and a brief discussion of price under conditions of imperfect competition. The part on distribution has been twice largely rewritten, especially as to the subjects of rent and interest. Increased attention has been given to the relation of rent to competing uses of land, and the theory of economic rent has been given a more generalized treatment. In the third edition the whole section on distribution was given its present position as Part III, in the first part of the book, which we consider its logical position. In the second edition the chapter on government industry was materially expanded in order to present a broader picture of the facts of public ownership in various departments of industry.

The third edition appeared in 1936, at a time when extraordinary changes were taking place in many departments of economic life, especially those more directly affected by government action. This revision therefore involved thorough rewriting of the chapters on money and banking, with appropriate attention to the analysis of irredeemable money. The sections relating to taxation and the business of government, the industrial organization, and inter-regional trade and exchange were revised to meet conditions of the day, involving rearrangement of topics, rewriting of many chapters, and the introduction of several new chapters on such topics as the farm problem, regulation of competitive industry, and contemporary problems of foreign exchange. In these sections attention was given to the many economic problems raised in America by the "New Deal."

In the preface of the third edition, three years ago, we took occasion to note a tendency, which appears still in evidence, to view the extraordinary events of the last decade as in some way casting suspicion upon the generally accepted principles of economics, with

the inference that these principles must now be thrown in the discard and replaced by a "new economics." The prosperous years of 1922-1929 were regarded by not a few persons as having ushered in a "new era," with poverty abolished and the human race lifted to a level of prosperity such as the mind of man had not before conceived. Economic science, casting doubts upon the validity of this roseate dream, was to be cast aside in favor of the "new economics" of perpetual prosperity. More recently the sad experiences of adversity have brought forth another "new economics," the economics of depression, rationalizing for example permanent unemployment and permanent government relief on something like the present scale. It is our belief that a less superficial view would disclose the prevailing body of economic principles as generally competent to explain the economic events of recent years, whether of prosperity or of depression. We are furthermore of the opinion that only in the light of such fundamental economic principles may there be any true analysis and judgment of the "New Deal" and other recent developments. It is in this spirit that the present revision has been prepared. While the formulation of basic economic principles has undergone little change, these principles find fresh illustration in a wealth of new facts and are called upon to elucidate the economic problems, new and old, which confront the world today.

The following are some of the more important changes in the present edition. The whole section on price has been reorganized and largely rewritten in order to incorporate the concept of imperfect competition as an integral part of the theory of price. Imperfect competition and pure competition, as well as monopoly, have been merged in a unified treatment, which we believe not only is an advance in theoretical accuracy but serves also to make the treatment more realistic and practical. In the section on money and banking, some clarification in the statement of principles has been found possible, and the material has been rearranged in such way as to avoid repetitions and merge the recent complicated changes in American monetary and banking legislation into the treatment of the subject as a whole, in a manner that was scarcely possible three years ago. Descriptions of European and Canadian

banking have been carefully revised and brought up to date. Recognition has been given to recent advances in the theory and practice of "managed money," stabilization funds, etc. The chapter on nationalism and the tariff has been enlarged by the introduction of some discussion of recent official restrictions upon international trade and the problem of reciprocal trade regulation, including Secretary Hull's trade agreement program. The discussion of labor problems in Part VIII has been revised to take account of recent governmental policies affecting wages, hours, conditions of work, and the relations of employers and employees, including a realistic discussion of social security legislation. A substantial amount of new material has been added to the chapters on radical reform proposals in Part IX, including an impartial description and appraisal of socialism, Russian communism, and Italian and German Fascism, combining historical and theoretical treatment. Finally the entire book has been brought up to date with respect to statistics and all statements of fact.

We make grateful acknowledgment of the invaluable assistance which we have derived from our colleagues in the Economics Department of Yale University. Though the actual task of authorship has fallen to the three undersigned, this book represents, especially in its first edition, the combined contributions, direct and indirect, of a large group of teachers who for many years have been closely associated in the task of teaching elementary economics to the students of Yale University. To all of these, past as well as present colleagues, we acknowledge our debt.

In particular we are pleased to record our indebtedness to Professor Irving Fisher, whose influence upon the authors, his former students, has been profound; likewise to Professor Clive Day, who has given valuable suggestions in the field of economic history, to Professor Winthrop M. Daniels for assistance in the preparation of the chapters on railroads, to Professor Ray B. Westerfield for contribution of valuable material on modern banking practice and on the subject of risk and insurance, to Dr. Benjamin P. Whitaker for aid in gathering material illustrating the principles of price, to Professors A. J. Nichol and R. M. Bissell, Jr., who contributed much to the last two revisions of the chapters on price, to Mr. C. H.

Whelden, Jr. for assembling materials on the "New Deal" and other recent developments in applied economics, to Professor Ralph C. Jones and Mr. J. H. G. Pierson for suggestions in the fields of accounting and monetary theory respectively, and to Professor I. R. Barnes for advice on the public regulation of business.

We are indebted also for helpful criticism and suggestions from a host of teachers who have used this book in other institutions, among whom special acknowledgment is due to Lieutenant Colonel Herman Beukema of the United States Military Academy and to Professor Clair W. Swonger of the University of New Hampshire, who has given the exercises a thorough overhauling for this edition. General acknowledgment is finally due the authors of the standard works on many subjects embraced within the broad field of general economics, upon whose writings we have leaned beyond the possibility of specific acknowledgment in every case.

FRED ROGERS FAIRCHILD
EDGAR STEVENSON FURNISS
NORMAN SYDNEY BUCK

NEW HAVEN, CONNECTICUT
April, 1939

CONTENTS

PART I. THE ECONOMIC ORGANIZATION

	PAGE
CHAPTER I. FOUNDATIONS OF THE SCIENCE OF ECONOMICS	3
1. Social sciences	
2. Human wants: (1) Number and variety. (2) Food, clothing, and shelter. (3) Other wants. (4) Intermediate goods. (5) No limit to the wants of man	
3. Nature, the provider	
4. Economics: (1) Definition. (2) Economics as a science	
5. Utility	
6. Wealth: (1) Definition. (2) Are human beings wealth? (3) How to identify wealth	
7. The problem of definition: (1) Necessity and freedom of definition. (2) Requirements of a scientific definition	
8. Income	
9. Costs	
10. Net income	
11. Property: (1) Definition. (2) Division of property. (3) Property in free persons. (4) Wealth and property. (5) Documents in evidence of property	
12. Wealth and income. Fund and flow	
13. Value: (1) A common denominator. (2) Transfer and exchange. (3) Barter and money exchange. (4) Value. (5) Price. (6) Value and price. (7) The unit of value	
14. Exercises	
CHAPTER II. THE FACTORS OF PRODUCTION	24
1. Production	
2. The factors of production	
3. Land	
4. Labor	
5. Capital: (1) Description. (2) Capital in production. (3) The accumulation of capital. Saving. (4) The form of capital investments. Invention	
6. The capitalist class	
7. The entrepreneur: (1) His function. (2) His reward	
8. Exercises	

	PAGE
CHAPTER III. INDUSTRIAL STAGES	40
1. Economic history	
2. The household system : (1) General description. (2) The manor	
3. The handicraft stage :	
(1) The towns	
(2) Merchant guilds	
(3) Craft guilds	
(4) Medieval trade	
(5) The town market	
(6) The medieval fair	
(7) Municipal regulation of commerce	
(8) Comparison with the household system : <i>a.</i> The beginning of specialization. <i>b.</i> The growth of freedom	
(9) Judgment of town economy	
4. The domestic system	
5. The factory system	
CHAPTER IV. THE INDUSTRIAL REVOLUTION	55
1. Nature of the industrial revolution	
2. Why the industrial revolution began in England	
3. Coal and iron	
4. The textiles : (1) Woollen and cotton industries. (2) Spinning. (3) Weaving	
5. Use of power and growth of factories	
6. Reluctance to establish factories	
7. Machines and tools	
8. Transportation : (1) Roads. (2) Canals. (3) Railways	
9. Summary of the effect of the Industrial Revolution. (1) Indus- trial progress. (2) Effect on the laborers	
10. The Industrial Revolution in America	
CHAPTER V. THE FORM OF THE BUSINESS UNIT	74
1. The individual proprietorship	
2. The partnership : (1) General description. (2) The limited part- nership	
3. The corporation	
(1) Modern growth	
(2) Nature of the corporation	
(3) Capital stock	
(4) Bonds and notes	
(5) Stocks and bonds compared	
(6) The property account of a corporation : <i>a.</i> Principal items of assets and liabilities. <i>b.</i> Losses and insolvency. <i>c.</i> The value of stock. <i>d.</i> Interest payments and dividends. <i>e.</i> Stock divi- dends. <i>f.</i> Watered stock	

- (7) The corporation's income account
- 4. Exercises

CHAPTER VI. DIVISION OF LABOR AND THE CONTROL OF PRODUCTION 98

- 1. Cooperation
- 2. Division of labor
- 3. Economic effects of division of labor
- 4. Division of labor and the extent of the market
- 5. Territorial division of labor
- 6. Division of labor a form of coöperation
- 7. The control of production
 - (1) Conscious and unconscious control
 - (2) Regulation by legal authority : *a.* Protection of the individual. *b.* Promotion of national welfare. *c.* Mercantilism
 - (3) Laissez faire : *a.* Reaction from government control. *b.* Economic freedom. *c.* The motive to productive effort. *d.* The rule of price
- 8. Taking stock of our economic organization
- 9. Shall the system be scrapped? (1) What the present system accomplishes. (2) The present conclusion

Suggestions for Further Reading on Part I 120

PART II. THE FORCES DETERMINING PRICE

CHAPTER VII. PRICE AND THE CONSUMER 127

- 1. Introduction to Part II
- 2. The role of price : (1) Price controls the material enjoyments of life. (2) Price from the producer's viewpoint. (3) A pecuniary system
- 3. The behavior of prices : (1) A bird's-eye view. (2) Complaints about prices. (3) Prices fixed by statute or custom. (4) Are sellers or buyers free to fix prices? (5) Is price a mystery?
- 4. Consumer's choices
- 5. Marginal utility : (1) The subjective side of utility. (2) The nature of utility. (3) Relative marginal utility. (4) Definition of marginal utility. (5) Graphical representation. A digression. (6) Meaning of marginal utility. (7) The law of marginal utility. (8) Marginal utility of durable goods. (9) Some apparent price anomalies resolved. (10) Abundance and scarcity
- 6. Exercises

CHAPTER VIII. DEMAND 153

- 1. Individual demand : (1) Marginal utility and individual demand. (2) Significance of this analysis. (3) How both parties to a trade

- can gain. (4) Things one cannot afford. (5) Real people. (6) Is all buying irrational? (7) Habit and utility
2. Total demand : (1) Individual demand and total demand. (2) Definition of demand. (3) Time and place of the demand schedule. (4) Limitations of this sort of analysis. (5) Demand and desire
 3. The law of demand : (1) General statement. (2) Special cases. (3) Explanation of the law of demand
 4. Elasticity of demand : (1) General statement. (2) Degrees of elasticity. (3) The limits of elasticity. (4) Examples of inelastic demand. (5) Examples of elastic demand. Substitution. (6) Necessities and luxuries
 5. Exercises

CHAPTER IX. DEMAND AS IT CONCERNS THE SELLER 176

1. The market : (1) Definition. (2) Free market defined. (3) Who sets the price? Three kinds of markets. (4) Naming either quantity or price
2. Competition
 - (1) Definition
 - (2) Competition and monopoly. Degrees of competition
 - (3) Pure competition : *a.* Market conditions. Definition. *b.* The individual seller meets perfect elasticity of demand. *c.* The seller has no "price problem"
 - (4) Imperfect competition : *a.* Inertia and ignorance. *b.* Individuality of products. *c.* Trade marks and special service. *d.* Small number of sellers. *e.* Demand not perfectly elastic. *f.* Definition of imperfect competition. *g.* The liberty of imperfect competition. *h.* Limitations upon the seller's liberty. *i.* Liberty and discretion in big business. *j.* Quality competition in place of price competition
3. Efforts to control demand : (1) Under pure and imperfect competition. (2) Advertising
4. The business man's interest in demand
5. Ascertaining the facts of demand
 - (1) Indicators of consumers' desires
 - (2) Indicators of purchasing power : *a.* Consumers' incomes. *b.* Movement of goods from producer to consumer. *c.* Production records. *d.* Indexes of general economic conditions
6. Exercises

CHAPTER X. THE SELLERS AND THEIR COSTS . . . 204

1. The different viewpoints of buyers and sellers
2. Production to order or for a general market
3. Immediate sales problem : (1) General statement. (2) Reservation price
4. Remote sales problem

- (1) Cost of production controls: *a.* Cost and the lower limit to price. *b.* The upper limit to price. *c.* Prices not dependent on costs of existing goods. *d.* The lingering influence of past costs. *e.* Future cost the normal control
- (2) Analysis of cost
 - a.* Fixed and variable costs
 - b.* Large scale savings in variable costs
 - c.* Expanding output and unit cost
 - d.* The limit to large scale enterprise
 - e.* Individual costs dependent on general conditions
 - f.* Marginal cost: (*a.*) In general. (*b.*) Marginal cost and average cost
 - g.* Cost schedules and cost curves
5. Exercises

CHAPTER XI. COST OF PRODUCTION AND PRODUCTION POLICY 229

1. Under pure competition: (1) Price above average cost. (2) Price below average total cost and above average variable cost. (3) Price below average variable cost. (4) Short run and long run. (5) Conclusion
2. Under imperfect competition
 - (1) The seller and the demand for his product
 - (2) Getting the maximum profit
 - (3) Marginal revenue
 - (4) Marginal revenue and marginal cost: *a.* In general. Price above average cost. *b.* Price below average cost. *c.* Conclusion
 - (5) Imperfect competition and the elasticity of demand
3. Imperfect competition and monopoly
4. Monopoly and the public interest
5. Exercises

CHAPTER XII. SUPPLY 250

1. Introduction
2. Individual supply: (1) Normal condition. (2) An exception
3. Total supply
 - (1) Definition
 - (2) Supply and stock
 - (3) Supply pertains to pure competition
 - (4) Significance of the concept of supply
 - (5) Relation of quantity to price
 - a.* Long and short run
 - b.* Increasing costs
 - c.* Constant costs
 - d.* Elasticity of supply

- e.* No supply curve slopes downward
 - f.* The law of supply
 - g.* Long run and short run again
- (6) Further analysis of the nature of supply
- 4. Decreasing cost industries
- 5. The business man's interest in the supply side of the market:
 - (1) In general. (2) Measurement of potential supply. (3) Normal and residual changes. (4) Indexes of potential supply. (5) Seeking control of supply
- 6. Exercises

CHAPTER XIII. PRICE UNDER PURE COMPETITION . 270

- 1. Pure and imperfect competition
- 2. Determination of price and quantity exchanged. Equilibrium:
 - (1) Price determined by demand and supply. (2) Determination of the quantity exchanged. Clearing the market. (3) A free market. (4) The laws of demand and supply. (5) Market price and equilibrium price. (6) The law of price
- 3. Alternative concepts of demand and supply: (1) Two possible meanings of demand. (2) Choice of a definition of demand. (3) Two possible meanings of supply. Choice of a definition. (4) The importance of consistency
- 4. The causes of price change: (1) In general. (2) A change in demand. (3) A change in supply. (4) The retail egg market. (5) The effect of perfect inelasticity of demand or supply. (6) The effect of perfect elasticity of demand or supply. (7) The law of price change. (8) When demand and supply both change. (9) Practical illustrations
- 5. Decreasing costs
- 6. Appraisal of foregoing methods of investigation: (1) Theory and practice. (2) Market prices and imperfections of the market. (3) Economic friction. (4) The laws of price on the stock exchange. (5) Shall we discard our theory?
- 7. Competition
- 8. Exercises

CHAPTER XIV. PRICE UNDER IMPERFECT COMPETITION 305

- 1. Introduction
- 2. The role of substitutes: (1) Their significance. (2) Nature of substitutes. (3) Influence of substitutes on the size of demand. (4) Influence of substitutes on the elasticity of demand. (5) Conclusion
- 3. Interaction of sellers' policies
- 4. Price determination
 - (1) Long run and short run
 - (2) Price determination in the short run
 - a.* Price cutting

- b.* Limitations: (*a*) First limiting factor. (*b*) Second limiting factor. (*c*) Third limiting factor
 - c.* Conclusion
- (3) Price determination in the long run
- 5. Conclusions
- 6. Exercises

CHAPTER XV. INTERRELATIONSHIP OF PRICES. PRICE THEORY IN PRACTICE 324

- 1. Interrelationship of prices
 - 2. The present value of future income: (1) Capital and income. (2) Property and capital. (3) Property valuation. (4) Present worth. (5) Present goods more desirable than future goods. (6) Interest. (7) The link between future and present values. (8) The discount process. (9) Capital and income. (10) Discount and the laws of price. (11) Discounted income and cost of production
 - 3. Prices of the same goods in different markets
 - 4. Prices of competing goods
 - 5. Prices of complementary goods
 - 6. Prices of competing cost goods
 - 7. Prices of joint cost goods
 - 8. Prices of tandem goods
 - 9. Price theory in practice: (1) Practical aspects of the theory of value. (2) The overworked term "value"
 - 10. Exercises
- Suggestions for Further Reading on Part II 347

PART III. THE DISTRIBUTION OF WEALTH AND INCOME

CHAPTER XVI. ECONOMIC RENT 351

- 1. The meaning of distribution
- 2. Nature and definition of economic rent
- 3. Diminishing returns: (1) In agriculture. (2) The general principle of diminishing returns. (3) What is implied in diminishing returns
- 4. Diminishing returns translated into increasing marginal cost
- 5. Marginal cost equated to market price
- 6. The concept of producer's surplus
- 7. Contractual economic rent equivalent to producer's surplus
- 8. Lands of different quality
 - (1) Cultivation of lands of different productivity
 - (2) Economic rent on lands of different quality
 - (3) Intensive and extensive margins

(4) Urban lands: <i>a.</i> Inequality. <i>b.</i> Diminishing returns from urban lands	
9. Various qualities, uses, and rents of land	
10. The value of land	
11. Relation of rent to prices	
12. Quasi-rent	
13. Economic rent distinguished from other income	
14. "Unearned increment"	
15. Attacks on economic rent	
16. Exercises	
 CHAPTER XVII. INTEREST.	377
1. Time preference	
2. Time preference and interest: (1) Production of capital requires time. (2) Interest a payment for waiting. (3) The universality of interest	
3. Nature of the loan contract	
4. The loan market and the rate of interest	
(1) Introduction. Definition of rate of interest	
(2) The supply of loans: <i>a.</i> Reactions of individual lenders. <i>b.</i> Corporate saving	
(3) The demand for loans	
<i>a.</i> Borrowing for production	
(<i>a</i>) Purpose of the borrower	
(<i>b</i>) The productivity of capital	
<i>b.</i> Borrowing for consumption: (<i>a</i>) General description. (<i>b</i>) Consumption borrowing determined by time preference. (<i>c</i>) Borrowing of governments	
<i>c.</i> Total demand for loans	
(4) Determination of the market rate of interest	
(5) Adjustment of the individual to the loan market	
5. The rate of interest and the quantity of money: (1) Traditional fallacy. (2) Bank reserves. (3) Transition effects	
6. Lack of uniformity in interest rates	
7. True and nominal interest rates	
8. Changes in the rate of interest	
9. The capitalist class	
10. Opposition to interest as a private income	
11. Regulation of interest: (1) Early experience. (2) The modern field for regulation. (3) Futile and injurious interference with the interest rate	

CHAPTER XVIII. THE GENERAL LAW OF WAGES

407

1. The rewards of human exertion
2. The problem of the rate of wages

	PAGE
3. The demand for labor: (1) Why employers hire laborers. (2) Mar- ginal productivity of labor. (3) Productivity discounted	
4. The supply of labor	
5. The general law of wages	
6. Practical qualifications	
(1) Theoretical precision lacking	
(2) Interdependence of different types of labor	
(3) The sphere of bargaining: <i>a.</i> In general. <i>b.</i> Wages above marginal product. <i>c.</i> Wages below marginal product	
(4) Differences in wage rates within occupational groups	
(5) Trade unionism and wages	
(6) Systems of wage payment	
<i>a.</i> The time base	
<i>b.</i> Piece rate wages: (<i>a.</i>) General description. (<i>b.</i>) Piece rates and the economic law of wages	
(7) Local variations in wage rates	
7. Discounted marginal productivity is fundamental	
8. Other types of labor income	
9. Exercises	

CHAPTER XIX. DIFFERENCES IN WAGES. WAGE PROB-
LEMS 429

1. Differences in wages	
(1) Social stratification of the laborers	
(2) Wages of each group tend toward uniformity	
(3) Least uniformity in the highest group	
(4) Wages and the contrast of social classes	
(5) Permanence of cleavage between social classes: <i>a.</i> The facts. <i>b.</i> Explanation. Nature and nurture theories	
2. Real wages	
3. Changes in wages	
(1) The law of proportionality	
(2) Labor and the law of proportionality	
(3) Historical changes in the rate of wages	
4. Wages and the standard of living	
(1) The worker's theory	
(2) The true effect of the standard of living	
5. Wage regulation	
(1) By action of employers	
(2) Minimum wage laws: <i>a.</i> General summary. <i>b.</i> The United States fair labor standards act of 1938. <i>c.</i> Appraisal of mini- mum wage legislation	

CHAPTER XX. PROFITS 452

1. The accountant's profit statement	
2. Implicit rent, interest, and wages of management	

3. Definition of profits
4. Actual profits and losses : (1) Actual profits. (2) Losses. (3) The balance of profit and loss
5. The risk-bearing function : (1) Business risks. (2) Risk bearing. (3) The promoters and their profits. (4) The small-scale risk bearers. (5) Risk bearing not always a productive service
6. Profits tend to be short-lived
7. Profits not a price determinant
8. Criticism of the profit system
9. Exercises

CHAPTER XXI. PERSONAL DISTRIBUTION 467

1. Functional shares in distribution. Summary
2. Total and average national incomes
3. Comparative national incomes
4. Income distribution in America
5. Inequality : (1) In America. (2) Inequality universal
6. Personal distribution of wealth : (1) The evidence. (2) In America. (3) In other countries
7. Labor's share of the national income
8. Inter-relation of wealth and income distribution
9. Is inequality increasing? : (1) The question. (2) In America. (3) Evidence from other countries
10. The problem of "poverty"
(1) Conditions. (2) The family unit. (3) The "standard of living." (4) Family incomes. (5) The facts
11. Causes of inequality
12. The problem of inequality
13. Reform of inequality : (1) Charitable relief. (2) Social insurance. (3) Public education. (4) Limitation of personal and property rights
14. Exercises

Suggestions for Further Reading on Part III 490

PART IV. MONEY AND BANKING

CHAPTER XXII. MONEY. GENERAL PRINCIPLES. STANDARD MONEY SYSTEMS 495

1. Nature of money : (1) Examples of primitive money. (2) The common characteristic. (3) Money is local. (4) Definition of money
2. Functions of money : (1) Medium of exchange. (2) Standard of value. (3) Store of value
3. Government control of money : (1) In general. (2) Legal tender
4. Money and currency

	PAGE
5. Varieties of money and monetary systems (definitions and general description)	
6. Standard money systems	
(1) Principal kinds of money	
(2) The ideal money commodity	
(3) Types of standard money systems	
(4) The gold standard : <i>a.</i> The monetary unit. <i>b.</i> Coinage. <i>c.</i> Free coinage. <i>d.</i> The value relation between gold and money	
(5) The bimetallic standard	
<i>a.</i> General description	
<i>b.</i> When the two ratios are equal	
<i>c.</i> When market ratio is less than mint ratio	
<i>d.</i> When the market ratio is the greater	
<i>e.</i> Gresham's law	
<i>f.</i> Is a discrepancy between the ratios self-correcting?	
<i>g.</i> Bimetallism in the United States : (<i>a.</i>) Introduction. Early American money. (<i>b.</i>) Original coinage ratio. (<i>c.</i>) The coinage ratio changed. (<i>d.</i>) The end of bimetallism. (<i>e.</i>) Change in the market ratio. Conclusion	
<i>h.</i> Bimetallism in Europe	
<i>i.</i> Triumph of the gold standard	
<i>j.</i> The "limping standard"	
(<i>a.</i>) In general	
(<i>b.</i>) American silver dollars after 1873 : To 1878. Bland-Allison and Sherman acts. The Pittman act	
(6) The gold bullion standard	
(7) The gold exchange standard	
7. Exercises	

CHAPTER XXIII. IRREDEEMABLE MONEY SYSTEMS . 522

1. Origin and general features
2. American experience to 1879 : (1) Before the Civil War. (2) The greenback régime
3. Irredeemable paper money since 1914
 - (1) General desertion of the gold standard
 - (2) The present British monetary system
 - (3) Recent American monetary experience
 - a.* 1914-1933
 - b.* Bank holiday and gold embargo of March, 1933
 - c.* Official abandonment of the gold standard
 - d.* Devaluation of the dollar
 - e.* Gold reserve act and proclamation of 1934
 - f.* New monetary unit. Legal tender
 - g.* "Gold profit" and stabilization fund
 - h.* Silver policy : (*a.*) The facts. (*b.*) Monetary significance of silver. (*c.*) Political significance of silver

- i.* Summary. The present American monetary system
4. The irredeemable foreign exchange standard

CHAPTER XXIV. THE PRINCIPLES OF BANKING . . . 540

1. The nature of credit: (1) Financing production. (2) Credit
2. The commercial bank
3. Lending: (1) Example of borrowing. (2) Interest and discount.
(3) Proceeds. (4) Principal types of trade paper. (5) A bank's
balance sheet. (6) Discounting a note
4. A normal bank statement
5. Bank deposits
 - (1) Definition
 - (2) Discount and deposit
 - (3) The check
 - a.* Definition
 - b.* Payment by check
 - (*a*) Within the bank
 - (*b*) Collection between banks: In general. Clearing house.
Clearing for outside banks. The volume of clearings.
Out-of-town checks
6. Paying a loan
7. The normal business of banking
8. Interrelation of loans and discounts, deposits and notes, and cash
9. The reserve
 - (1) Its function
 - (2) Profits vs. safety
 - (3) How the reserve ratio is controlled: *a.* Buying and selling
securities. *b.* Rediscounting. *c.* Borrowing. *d.* Calling loans.
e. Control by the rate of discount
10. Note issue
 - (1) Introduction
 - (2) Notes and deposits: *a.* Fundamental similarity. *b.* Differences
between notes and deposits
 - (3) Legal restrictions
11. The problem of elasticity: (1) Need of elastic currency. (2) Elas-
ticity of bank credit. (3) Influence of reserves
12. Other banking and credit agencies: (1) The note broker. (2) Com-
mercial paper houses. (3) Trust companies. (4) Savings banks.
(5) Private banks
13. The function of credit: (1) Credit in general. (2) Peculiar impor-
tance of bank credit
14. Exercises

CHAPTER XXV. THE AMERICAN BANKING SYSTEM . . . 578

1. Historical background
2. Banking in the United States before the Civil War

	PAGE
3. The national banking system: (1) Origin. (2) National bank notes. (3) The problem of reserves	
4. The Federal Reserve	
(1) Organization: <i>a.</i> Board of Governors. <i>b.</i> Federal reserve districts and banks. <i>c.</i> Member banks	
(2) Functions of the Federal Reserve	
<i>a.</i> Rediscounting and discounting	
<i>b.</i> Federal reserve notes: (<i>a</i>) General description. (<i>b</i>) Elasticity of federal reserve notes. (<i>c</i>) Plan to retire the national bank notes. (<i>d</i>) End of the national bank notes	
<i>c.</i> Reserves against deposits	
<i>d.</i> Open-market operations	
<i>e.</i> Centralized reserves	
<i>f.</i> Mobile reserves	
<i>g.</i> A national system of clearing	
<i>h.</i> The government's financial business	
5. The American banking situation in 1929: (1) Introduction. (2) Composition of the banking system. (3) Branch banking and chain banking. (4) Weaknesses of decentralized banking. (5) Affiliates. (6) Expansion of credit. (7) Foreign loans. (8) Stimulation of exports. (9) Gold exchange standard	
6. The financial crises of 1929 and 1931	
7. Bank failures	

CHAPTER XXVI. AMERICAN BANKING SINCE 1929 . . . 609

1. Banking legislation since 1929	
2. Pre-"New Deal" remedies: (1) National Credit Corporation and Reconstruction Finance Corporation. (2) Glass-Steagall act	
3. Banking collapse (in 1932-33)	
4. Emergency banking legislation: (1) Original acts. (2) Perpetuation of emergency legislation	
5. Structural reforms of 1933 and 1935	
(1) Deposit insurance: <i>a.</i> Description. <i>b.</i> Extent of deposit insurance. <i>c.</i> Judgment of deposit insurance	
(2) Capital requirements	
(3) Branch banking	
(4) Holding company affiliates	
(5) Divorce of commercial and investment banking	
6. Functional remedies of 1933 and 1935	
(1) Interest on deposits	
(2) Speculative uses of credit	
(3) Centralized control of the banking system	
(4) Control of credit: <i>a.</i> Concept of credit control. <i>b.</i> Open-market operations. <i>c.</i> Regulation of reserve ratios of member banks. <i>d.</i> Loans on real estate	

7. The present banking situation : (1) In general. (2) Statement of the federal reserve banks
8. Other forms of government banking : (1) Agricultural credit agencies. (2) Loans to home owners
9. Criticism of government banking

CHAPTER XXVII. FOREIGN BANKING SYSTEMS . . . 636

1. Banking principles and banking systems
2. The French banking system
 - (1) Contact with the public
 - (2) The Bank of France : *a.* Ownership and control. *b.* Character of business
3. British banking
 - (1) Commercial banks
 - (2) The Bank of England
 - a.* Character of business
 - b.* Concentration of reserves
 - c.* Note issue
 - (*a*) General description. Issue Department
 - (*b*) Elasticity of notes : Elasticity sacrificed. How Great Britain secures elasticity : Ordinary mechanism. Extreme procedure in time of crisis
 - d.* The Banking Department
 - e.* The reserve
4. The Canadian banking system : (1) The "chartered banks" and their resources. (2) Notes. (3) Record and judgment of Canadian banking. (4) The Bank of Canada

CHAPTER XXVIII. MONEY AND PRICES . . . 654

1. The value of money
2. Value of money not constant
3. Measuring the value of money
4. Index numbers : (1) Price relatives and index numbers. (2) The average, simple and weighted. (3) An index number of the aggregative type. (4) Practical index numbers
5. The problem of the price level
 - (1) Introduction
 - (2) Volume of trade
 - (3) Quantity of currency
 - (4) Velocity of circulation
 - (5) The quantity theory of money : *a.* Algebraic statement. *b.* Significance of the quantity theory. *c.* Practical limitations of the quantity theory
 - (6) Historical evidence from American experience : *a.* Summary of price movements. *b.* An increase in the quantity of money,

CONTENTS

xxvii

PAGE

irredeemable money (1861-1864). *c.* A decrease in the quantity of money (1864-1868). *d.* An increase in trade (1868-1879). *e.* Increases in money and trade (1879-1897). *f.* Monetary inflation through gold. *g.* The world's production of gold. *h.* Effect on the price level (1897-1913). *i.* Monetary inflation of the World War (1913-1920). *j.* Lower prices since 1920. *k.* Irredeemable money and prices (1933-1938)

6. Far-reaching effects of monetary fluctuations

7. Monetary reform

8. Managed money

9. Exercises

Suggestions for Further Reading on Part IV 689

INDEX 693

PART I
THE ECONOMIC ORGANIZATION

I

FOUNDATIONS OF THE SCIENCE OF ECONOMICS

Social sciences. Among the various sciences which the pursuit of human knowledge has developed, there are those whose subject matter includes man himself, and among these are certain sciences concerned with the study of human beings, not as isolated specimens, as in physiology or human anatomy, but as members of social groups — families, tribes, villages, states, nations. The sciences which have to do with man in his various relations with his fellow men are known as the social sciences, among which economics holds a prominent place.

The particular province of economics, which distinguishes it from the other social sciences, is derived from two of the most fundamental principles of nature, one having to do with the character of man himself, the other relating to the natural environment in which man lives.

Human wants. Man is an animal, one of the millions of species of living things — plant and animal — that populate the earth. Nature has brought man and all these other forms of life to their present state as separate species and to their present equilibrium among themselves as sharers of the earth, through countless ages of evolution. The history of this process is a fascinating story, but one which enters only incidentally into the study of economics. Our task is to inquire into the nature of man as we find him in the world of today, without neglecting to recognize the fact that he is the product of a process of evolution, which is doubtless still going on, though so slowly that we are justified in considering man's present position in nature as fixed for all practical purposes.

The characteristic of man which most concerns our present inquiry is the infinite number and variety of his wants, and the first stage in marking out the field of economics is to acquaint ourselves with the nature and peculiarities of human wants. For the first

step or two we find nothing to distinguish man from the lower animals. In common with all species of animal life, man has certain organic needs which must be met or he will die. He must have food and water; he must have air to breathe. His bodily temperature must be always maintained somewhere between 98 and 99 degrees, and the remarkable thermostatic mechanism which nature has provided him for this purpose cannot cope with excessive exposure to heat or cold. Hence man requires certain forms of clothing and shelter, depending upon the climate. Man is also subject to attack by natural foes, and he must therefore have the means wherewith to defend himself or to escape. The simplest means appropriate to satisfy these fundamental organic needs are called "necessities," in the narrowest sense of that somewhat ambiguous term.

But now we come to a sharp distinction between man and the lower animals. The latter are satisfied when the necessities of life are met. With man the necessities are only the beginning. No sooner are they met than he reaches out for the satisfaction of other wants, for the "comforts and luxuries." Even the means whereby modern men satisfy the fundamental organic needs have passed far beyond the realm of bare necessities. This attribute of man is so fundamental to every phase of human life, and in particular to man's economic activities, that we must take pains to grasp its full import even at the risk of what may at first sight appear as an elaboration of the obvious.

Food, clothing, and shelter. Food is necessary to prevent starvation of course, but that is not what the average person is thinking about as he seats himself at the dinner table. He is thinking of the joy of eating. He studies the menu. He rejects oysters and decides on clams. He scans the soup list. He ponders on the choice of roasts: chicken, turkey, beef, lamb; and so on down through the vegetables, salad, dessert, fruit, and coffee. This ceremony is something more than the warding off of starvation. And the same is true, only in lesser degree, of the simplest home table. To the original hunger instinct has been added a host of tastes and desires, so that we demand of food a good deal more than that it keep us alive. The strong desires that have been imposed

upon the foundation of the drink-seeking instinct are too well known to need elaboration. Drinking is certainly not confined to the supply of moisture to keep the body alive.

The necessity of protection from heat and cold gives rise to the desire for clothing. But to the modern civilized man or woman clothing means a good deal more than this. In buying a suit far more attention is given to the pattern, the cut, the style, the fit of the garments than to their cold-resisting qualities. Hats, coats, dresses, shoes, stockings, collars, ties, and so on down the list — how much of all this could be eliminated if all we wanted were protection against the elements! In fact the demand of civilized man for clothing is fully as much a response to vanity, and it is well known that in climates where clothing is not a necessity such clothing as is worn by the savages is often the result of the desire for ornament to satisfy the wearer's vanity.

In most climates some sort of shelter is a necessity. The Eskimo with his snow house and the tropical savage with his palm leaf thatch on a bamboo frame are catering to this elementary want and not much more. But the dwellings of civilized men, from ancient days down to the present, have been much more than this. From the humble cottage to the royal palace, men have sought to make of their dwellings, not mere shelters from the weather, but things of beauty, of comfort, of ostentatious display, of luxurious enjoyment. One will generally have to search long in the modern village or city to find any habitation that does not represent a craving for something beyond mere shelter. Our houses, like our clothes, are responsive to vanity. They also represent (often grotesquely enough, to be sure) our aesthetic desire for that which is beautiful.

Other wants. Thus we have commenced to list the wants of mankind. We might continue, mentioning for example some of the things that the need of defense requires: strongly built houses, locks on doors and windows, protective weapons, policemen, armies and navies and their equipment, drugs and medicines, hospitals, jails and insane asylums, insurance companies — this is simply the beginning of the list. Consideration of the gregarious tendencies of mankind reminds us of the concentration of people in cities,

towns, and villages, thus giving rise to special wants in the field of housing, transportation, etc. Even when people seek relaxation from their ordinary activities and surroundings and take vacations "in the country," they do not generally seek solitude. They require seashore and mountain hotels, resorts, and camps, where they can relax in company with others. Man has a thirst for knowledge, in its elements perhaps instinctive, but carrying him far beyond the promptings of mere instinctive curiosity. He wants books, telescopes, laboratories, schools, and colleges. He wants to travel and to visit new and strange scenes and so demands transportation facilities, hotels, and guides. We have a desire for the beautiful and so want music and works of art. We enjoy sport and recreation and therefore require golf clubs, baseball outfits, swimming suits, and all the paraphernalia of modern sport.

Intermediate goods. We should also note here the fact that modern civilized man wants many things which are not capable of satisfying directly any desire. The housewife wants a sewing machine, though there is obviously no satisfaction directly obtainable from the possession or use of this article. But the sewing machine will serve as an instrument for making and repairing the family clothing, thus helping to satisfy a fundamental need. It is not wanted for itself, but as a means to the gratification of another need. This is a characteristic of the whole group of things known as instruments, tools, and machines, which make up a surprisingly large part of the commodities which men want and for which they strive. They are called "intermediate goods" and will demand further study in the next chapter and elsewhere in our investigation.

No limit to the wants of man. This sort of study might be carried to any length. Enough has been said however to bring out the important conclusion toward which we have been aiming; namely, the infinite number and the bewildering variety of the things that man wants. There is actually and literally no limit. Satisfying the instinctive cravings, securing the things necessary to sustain life and perpetuate the race — this is only the barest beginning. We soon forget that sustaining life is the basis of any of our wants. We are never content. The more wants we satisfy, the wider is the horizon opening up new and tempting vistas of

wants still to be catered to. We are quite justified in our conclusion that man's wants, as regards their number and variety, are without limit. This is the first of the two great principles which mark out the field of economics.

Nature, the provider. The second principle is that the where-withal to satisfy man's unlimited wants does not come without effort on his part. Nature may be a bountiful provider; she is in fact the source of everything man has or can hope to have. Some of the things he wants she provides gratuitously. We have plenty of air to breathe; we have sunlight, though not always in the exact amount we might like; water is generally, though by no means always, furnished in abundance by nature. A certain amount of food, shelter, and clothing may sometimes be obtained from nature with little or no human effort, though this holds true only of the simplest wants of the most primitive men and even then only in exceptional cases. But after noting these few exceptions we must conclude that nature does not generally bestow her gifts with a free hand; practically none of the infinite number of things that man wants is furnished him freely by nature. Man must work for what he gets. Even the lower animals have to exert themselves to provide for their needs, and if man had been content with the sort of living which the lower animals thus obtain he would still be a lower animal himself. Man is always engaged in efforts to secure the things he needs for the satisfaction of his wants. His wants are unlimited; the things to satisfy them are strictly limited. There is never enough to go around. Here we have the explanation of the greater part of the activities of human beings through all history. Mankind has always been chiefly engaged in "working for his living." Here we have also the roots of the great human institutions, law, property, the family, etc., which have developed out of man's struggle for the things with which to satisfy his wants.

Economics: Definition. The insatiability of man and the niggardliness of nature are thus the foundation stones upon which rests the structure of economics. Starting with these two premises, economics investigates the activities of man in his efforts to satisfy, so far as he may, his limitless wants out of the limited resources granted him by nature. We have then this preliminary and formal

answer to the question: what is economics? *Economics is the science of man's activities devoted to obtaining the material means for the satisfaction of his wants.* Lest the reader miss the true inclusiveness of this definition, he is reminded of what will be made clearer a little later; *i.e.*, that the "material means" which serve to satisfy human wants include human beings as well as external objects.

When we stop to consider that an important part, not to say the greater part, of all the waking hours of most adult men and women is devoted in some form or other to the business of acquiring the means for satisfying wants, and that even the children are much engaged either in this same endeavor or in acquiring the education and training necessary to equip them to begin it a few years hence, we can realize that, although strictly limited by the definition, the field of economics is a broad one. We might proceed to catalogue the many topics included and the various problems to be investigated. But a mere glance at the table of contents of this book will suffice to give the reader a general notion of the subject matter of economics. We may let this stand as the preliminary answer to the demand for a notion of the content of economics and the reason for its investigation. The complete answer may be expected to unfold itself as our inquiry proceeds.

Economics is a science. We have been speaking of the science of economics, and it may help the reader to a still clearer notion of what economics is and also enable him to avoid some seductive pitfalls as he further pursues his path of investigation if we pause here to inquire briefly into the nature of science. A science may be defined as a body of knowledge in some particular field arranged in an orderly way to show relations of cause and effect. Science deals only with facts and causal relations. It is not the business of science to tell how things ought to be, to lay down moral precepts, to advocate that which is good, or to condemn that which is evil. Nor is it the business of science to lay down rules for the accomplishing of desired ends. That is the function of an art. Thus the sciences of astronomy and mathematics set forth certain facts and causal relations; these are employed by the art of navigation to accomplish its desired ends. The sciences of physics and biology are similarly availed of by the art of medicine.

In thus sharply marking out the field of science we do not imply that moral or ethical considerations are to be ignored by the student of economics. Economics is a social science, and as such it grasps as its subject matter all attributes of man which affect his pursuit of the material means for the satisfaction of his wants. Among these attributes the moral and ethical notions of mankind are of importance and require full consideration. For example the moral code of one community may dictate polygamy, while that of another insists upon monogamy, and these moral edicts are of great economic significance. The science of economics will tell us a great deal about the respective effects of these two systems of family organization, but we must not forget that economics cannot and does not attempt to tell whether monogamy or polygamy is "right or wrong."¹

Utility. We need now to acquire clear notions of certain fundamental concepts which are inherent in the subject matter of economics and are an essential part of the equipment of the student.

Man desires those things that are able to satisfy his wants. Anything that satisfies a human want is useful; it has utility. *The utility of a thing is its capacity to satisfy a human want.* This definition scarcely needs elaboration or explanation. We determine readily enough whether we want any particular thing; *i.e.*, whether it is "useful" or "useless."

In consequence of its character as a science, economics gives somewhat broader meanings to "useful" and "utility" than are employed in common speech. When we call anything useful, we imply nothing as to its merits in an ethical sense. A thing is useful if it can satisfy a human want; that is all. We do not inquire whether the satisfaction of that particular want may be good or bad, wise or foolish. Opium, except when used as a medicine, may be physically injurious and morally bad, but this does not alter its classification as an object of utility so long as it does actually enable a human want to be satisfied. The use of chewing gum is probably foolish, but we cannot thereby deny its utility. The pictures which adorn the walls of many homes may be wholly bad from an artistic sense, but they are nevertheless economically useful. Of

¹ Cf. Francis A. Walker, *Political Economy*, Third Edition, pp. 19-21.

course this does not imply any lack of appreciation of the importance of aesthetic and ethical judgments. Such questions are simply outside the field of economics, as of any other science.

It is important to realize also that the utility of a thing is not a given quantity dependent only on the physical characteristics of the thing itself. It depends on where the object is found, on who wants it, and on the availability of substitutes. A leg of lamb in a packing house at a distant city has less utility than the same leg of lamb in the local butcher's cooler. Further the same leg of lamb has much greater utility for Mr. X, who is very fond of lamb, than for Mr. Y, who has never been able to eat it. Finally if Mr. X is presented with a large and delectable rib roast of beef as a gift, the leg of lamb will possess less utility for him than if he had no other meat at hand.

Wealth: Definition. If from the category of useful material things we eliminate those few which nature furnishes in abundance sufficient to satisfy man's wants, we shall approach what the economist calls "wealth." The science of economics is largely built up about wealth, so much so that economics has often been defined as "the science of wealth," a definition which, properly understood, is almost synonymous with that which we have accepted. It is indispensable therefore to have a clear idea and a precise definition of wealth.

Wealth consists of all useful material things owned by human beings. This definition needs some further examination. We have already settled the meaning of "useful." Useless things are not included in wealth, which thus embraces only those things which are capable of satisfying human wants. The reader should perhaps again be warned that the ethical character of human wants does not enter into the classification.

We limit wealth secondly to material things. This excludes such desirable things as honesty, good health, the speed of a race horse, the coldness of ice, the skill of an artist or a musician. This may possibly seem arbitrary and unfortunate at first. But we shall not have to look far to see the great advantage of this limitation and particularly the confusion which would follow if the concept of wealth were not thus limited. If the speed of a great race horse

is wealth, what of the horse himself? Certainly he is wealth. But have we then two articles of wealth, the horse and his speed? And if so, should we not add a third, his strength; and a fourth, his beauty? Most of the forms of immaterial wealth with which we should have to deal if our definition admitted them are simply qualities of material things which are themselves wealth. To count both the material thing and its qualities as wealth would be the crudest sort of double counting and productive of nothing but confusion. This is really to confuse wealth and utility, since most of the examples of immaterial wealth which would be urged are simply those qualities of some material object which enable it to satisfy a human want. Similar confusion would arise from including human qualities, such as health, strength, skill, honesty, in the category of wealth. We shall consider a little later the question whether human beings themselves are wealth.

Having excluded from our concept of wealth all useless and all immaterial things, we next limit it to those useful, material things which are owned by man. The first effect of this limitation is to exclude those useful things which are furnished freely and abundantly by nature. Thus air is material, and it is about the most useful thing in the world. But on account of its abundance it is not owned. Why should one take the trouble to own it? We also exclude certain things which, from their nature or from the nature of the human institution of ownership, cannot be owned. We cannot own the sun or the moon; no person or group of persons can own the Atlantic Ocean or the Gulf Stream. Most material things that satisfy human wants however are owned. Ownership may be vested in the individual person, as in the case of one's suit of clothes, the farmer's land, the merchant's store, etc. Or various groups of persons, such as partnerships, corporations, etc., may own land and buildings and virtually any kind of wealth which may be owned by individuals. Governments — national, state, and local — are large owners of wealth; they are simply another form of association of human beings.

Are human beings wealth? The requirement of ownership by human beings restricts the scientific concept of wealth in another important respect. Are human beings themselves wealth? As

regards slaves, the question is clearly answered in the affirmative by reference to our definition. The slave is useful; he is a material thing; he is owned by his master. But how about free men or women or children? They are not owned under the legal systems of most modern nations. We might set up the fiction that free persons are owned by themselves or (in the case of minor children) by their parents. Free persons would then fall into our category of wealth, and there are some advantages to this classification. Free persons are material things of course, and they have qualities which enable them to satisfy wants of other persons. Moreover it is not always easy to know just where to draw the line between free men and those who are owned. On the other hand, it is not customary to think of free persons as owned, even by themselves, nor do we naturally think of free persons as wealth. The weight of advantage is on the side of our definition as stated, according to which, while slaves are undoubtedly wealth, free persons are not. Throughout the whole study of economics however it will be found necessary to take account of the utility of free persons and their services in satisfying human wants on the same terms as the utility and services of wealth.

How to identify wealth. We thus arrive at our definition of wealth. To determine whether any given thing is wealth or not, we have only to ask three questions: is it useful? is it material? is it owned? If the answer is yes to all these questions, the thing is wealth. A negative answer to any one of the questions excludes it.

The problem of definition: Necessity and freedom of definition. The foregoing discussion, leading up to a precise definition of wealth, may possibly appear unnecessary and arbitrary. Do we need a precise definition of such a common word as wealth? And if so, have we any right thus to set up a definition to suit ourselves? In answer to the first question, we need to remind ourselves that agreement upon precise definitions of all technical terms is one of the first requirements of science. Economists have not always realized the importance of this requirement or that it applies to their own science no less than to others. It is sometimes assumed that, since the technical terms of economics are largely those of everyday speech and writing, everyone must have a clear

enough knowledge of their meanings. It is agreed that the physicist must define "ampere" and "volt," because no one would otherwise know what he was talking about. But why should the economist have to define "wealth," "income," "money," and other common terms? Doesn't everyone understand well enough what these words mean?

Now the fact is that everyone does not understand well enough what these terms mean. Ask a dozen intelligent persons to define wealth; you will probably get a dozen different definitions. Is sunshine wealth? Is a burglar's jimmy wealth? How about the moon, the skill of a great surgeon, the town hall, the month of January, the Gulf Stream, the honesty of a bank cashier, etc.? People may think they have clear ideas of the meanings of common words until they are asked to formulate and defend a definition; then the vagueness of popular usage becomes evident. Such vague ideas may do well enough for ordinary use, but science requires precision. Economics therefore cannot evade the obligation of defining its technical terms, even though they be words of everyday use.

Indeed the adoption of such common terms makes the problem of definition harder rather than easier. Everybody will admit that there must be a definition of "ampere" and will see that there is nothing to do but learn and use the definition exactly as given. But economics must always contend against a certain feeling that definitions of common terms are not really necessary and against the vague and contradictory preconceptions which have become attached to the words of everyday use. This difficulty must be met squarely; it cannot be evaded.

Having accepted the obligation to define our technical terms, we revert to the question: are we free to set up our definitions as we see fit? The answer is yes. Names and definitions are to a certain extent arbitrary. There is no absolute test of the correctness of a definition; it is idle to argue about it. The reader who is familiar with Mark Twain's compilation of *Extracts from Adam's Diary* will recall the frequent disagreements that arose between Adam and Eve over the naming of the animals in the Garden of Eden, a task which Adam found sorely taxing his powers but which was

simplicity itself to Eve. When for the first time there appeared the dodo, Adam was at a loss for its name. Eve announced the correct name at once and, to all demands for proof or explanation, replied scornfully, "the moment one looks at it one sees at a glance that it looks like a dodo." So it became the dodo. If in the development of the English language the term "dog" had become attached to the biped now known as a duck and the term "duck" had similarly become associated with our favorite quadruped friend, we should get along just as well as we do now. Words have meanings which have become attached to them through the long process of the evolution of language. To argue whether these meanings are correct or not would be a foolish waste of breath. And little is to be gained by an argument over the "correct" definition of wealth.

Requirements of a scientific definition. On the other hand, definitions are not a matter of indifference. In science much depends upon the choice of terms and their careful definition. While the test of absolute correctness is wanting, there are two simple pragmatic requirements to which every scientific definition should conform. First the definition must be useful for the purposes of scientific study, and second it should agree as closely as possible with popular usage. These two requirements may come into conflict with each other, and in that case it is generally the second which must give way. This sacrifice is not so serious as might at first appear. Since popular usage is always more or less vague, a scientific definition must in any event differ somewhat from the popular definition or definitions. Exact agreement with popular use being out of the question anyway, the departure required to make the definition useful is generally not serious.

We justify our definition of wealth therefore on the ground that it will be useful in our further investigation without departing too far from common usage. That we have fairly well met the second of these tests will probably be apparent at once. The usefulness of our definition will be more fully demonstrated of course after the student has finished this book, or better still after he has continued his study of economics with the aid of other authors or through his own research.

Income. Wealth must be useful. When an article of wealth satisfies a want of man, we say that it furnishes *benefits* or renders *services*. By the benefits or services of wealth we mean *desirable events which it causes for human beings*. Keeping out the cold is a desirable event caused by a dwelling house; it is one element of the service of shelter or the benefit yielded by the house. The service of a piece of ice in the refrigerator is keeping things cold. A diamond ring renders service by satisfying the love of beauty or perhaps only the vanity of its wearer. The service of a slave is working in the field. Thus every article of wealth renders its appropriate service to man, else it would not be useful and so would not be wealth. In exactly the same way services are rendered by free persons, as when an actor performs before an audience or a lawyer tries a case for his client. The word *income* is used to include such services and is defined as follows: *Income consists of the benefits or services rendered by wealth or by free persons*.

Costs. All the events caused by wealth are not desirable. Besides furnishing shelter, a dwelling house requires painting and repairs, it compels its owner to pay taxes and insurance premiums, and so on; finally it wears out or becomes obsolete and if not discarded must be torn down and rebuilt. The slave works in the field, but he requires food, clothing, and shelter; and the free farm laborer requires wages. *The undesirable events caused by wealth are called the disservices or costs of wealth*.

Net income. Normally all wealth yields both income and costs. We cannot escape the undesirable events if we would have the desirable events; we put up with the costs in order to get the income of wealth. It follows that, in the long run, the income from any article of wealth must exceed its cost (how these are measured will appear later), else no one would care to own that article and it would cease to be wealth. For example a storekeeper has a motor truck to deliver goods to his customers. The truck is an article of wealth; the income from it is the service of delivering goods; its costs are the expenditures of money for the gasoline and oil it consumes, garage room, occasional washing, and repairs. These costs would not be borne if the service of delivering goods were not worth more than the costs. Eventually the truck begins to wear out;

its speed diminishes, and it cannot make some of the grades. It consumes more gasoline, and its loose piston rings and bearings are wasteful of oil. Repair bills become heavier and more frequent. Thus its income declines while its costs rise, and sooner or later the margin between them will disappear. The machine is no longer useful and it finds a resting place on the junk heap; it is no longer wealth. *The difference between the income and the costs of any article of wealth is its net income.* Some net income must be present or the article will not be wanted. In general then utility means the power to yield a net income or an income in excess of the costs, and this idea of a net income is inherent in the term "useful" as employed in the definition of wealth.¹

Property. Every article of wealth is owned by somebody. This relation between wealth and its owner is one of the most fundamental concepts of economics. It may be called ownership, property right, or property. *Property* is the technical term generally used in economics. It is defined as *the right to income; that is, the right to the benefits or services of wealth or free persons.* This evidently is what ownership means. When a person owns an article, he has a right in it; namely, the right to have it and use it, all of which is included in the right to the services or income from that article. The owner is protected in his property right by the laws and customs of the community in which he lives; other persons are prevented from interfering with his enjoyment of the services of his wealth. Since the term "income" is so broad as to include any use to which the owner could put his wealth — including even its destruction if he so desires — property might be defined simply as the right to wealth or free persons, without any change in meaning. The definition here chosen has the advantages, (1) of emphasizing the fundamental importance of income, (2) of somewhat readier adaptation to the innumerable cases of divided ownership, and (3) of avoiding any apparent conflict with the common usage of not regarding free persons as owned.

¹ The reader who has any familiarity with accounting will note that these definitions of income, costs, and net income differ somewhat from those used by the accountants. They are not really in conflict. The accountant's terminology represents a special technical application of the more general concepts with which we are here dealing.

Division of property. A sole owner has the exclusive right to all the income of the article of wealth in question, as in the case of the owner of a house and lot, an automobile, a suit of clothes, etc. Ownership may however be divided, in which case each of the two or more owners has certain rights in the wealth; that is, the right to certain of its benefits or income. The tenant of a rented house has the right to its use only during a certain time, as limited by his lease contract with the landlord, who has retained the right to the services of the house at all other times; that is, before the beginning and after the termination of the lease. In a partnership the wealth held belongs jointly to the two or more partners. Each has the right to a certain part of the services or income of this wealth, the exact division being specified in the partnership agreement. So in the corporate form of business organization, the wealth of the corporation belongs ultimately to the stockholders, of whom there may be hundreds or thousands, each with certain rights in the income from the corporation's property. The bondholders of the corporation represent another group of part owners, with certain other rights in the corporation's wealth, all specified in the bonds. Thus each bondholder has the right to receive certain interest payments at certain specified dates and finally at a specified date to receive a certain larger sum equal usually to the amount which was loaned to the corporation. That part of the income from the corporation's property which remains after the rights of bondholders and possibly other classes of part owners, such as note holders, etc., have been satisfied belongs to the stockholders and may be distributed to them in the form of dividends. Property may thus be divided in a great variety of ways. Every person who has a property right in an article of wealth is an owner (sole or part) of that wealth, though in common speech the term "owner" is sometimes applied to only one of the parties. Thus the landlord and the tenant are really both owners of the rented house. Neither is the exclusive owner, for neither has the right to all the services of the house. The property is divided between them, though we are accustomed to apply the term "owner" to the landlord alone. In like manner the bondholders are not usually spoken of as owners of the corporation, though fundamentally they are part owners.

All the services of wealth must belong to somebody. When property is divided, the sum of all the property rights of the several part owners must equal the total property right to all the services of that piece of wealth. For every article of wealth there must be a property right or property rights, and back of every property right there must always be an article of wealth or a free person or free persons.

Property in free persons. The last words of the previous paragraph will recall that property may be a right to the services of free persons as well as of wealth. If we had chosen to include free persons in the definition of wealth, all that has been said of property in wealth would have applied equally to them. As it is, we must recognize that free persons render services to others and that these others have rights to such services. A baseball player signs a contract with a manager agreeing to play exclusively for him during the coming season. The manager has thereby a property right, the right to receive certain services from this player. To that extent the player's right to his own time and services — actually his right to himself — is diminished. It is a case of divided property again. Theatrical managers, moving picture producers, musical promoters, and others have property rights in actors and actresses, musicians, and other artists. Every employer has a property right to certain services of his employees. A promissory note gives the creditor a property right in the debtor; that is, the right to receive a certain sum of money. Lawyers have rights against their clients, and clients against their lawyers; parents have rights in their children, and children in their parents. Every business contract gives each party a property right against the other. In fact the property in persons is almost always divided. Practically no one is completely his own master, and there are all degrees of property rights of others, from the slave, a true article of wealth because completely owned by his master, through serfs, peons, indentured servants, to the freest so-called "free man."

Wealth and property. In popular speech, the terms property and wealth are not carefully distinguished. It is quite usual for example to speak of a house and lot as "property," to call household furniture "personal property," and so on. For the purposes

of scientific analysis, economics draws a sharp distinction between property and wealth. The latter is a physical, tangible thing; it is material. Property, on the other hand, is a relation between wealth (or free persons) and persons; it is immaterial. This distinction is fundamental and has many important bearings. Thus wealth is the source of income; whereas property is the distributor of income. The wealth of a community, together with its productive free persons, determines how much and what kind of income the people may enjoy. The property rights of the several members of the community determine how these incomes shall be divided among them, how much of the total community income each one shall enjoy. Property rights may be changed without changing the total amount of wealth. The material welfare of the human race is a matter of the satisfaction of wants and is affected both by the total quantity of wealth and by the way the property rights to wealth are distributed among individuals, families, and nations.

Documents in evidence of property. Property rights are sometimes, though by no means always, expressed in legal documents. The owner of a farm or a building lot has a *deed*, which is a piece of paper containing a legal statement of his property right to all the benefits of a specified parcel of land. When one purchases the right to sit in a certain seat of a theatre on a certain date, his property is evidenced by a ticket. On the other hand, no legal document testifies to one's property right in the clothes he wears or the food on his table. Whether there shall be a document or not is a matter of custom or law. The property is the original and fundamental thing; the document, when there is one, is secondary and incidental, and in many cases no document is considered essential. The reader must be on his guard against confusion between property rights and the documents in evidence thereof, particularly in those cases where common, and even scientific, usage has not found it necessary to employ different terms for these two things. For example a bank note is, strictly speaking, a piece of paper testifying to a property right of the bearer against a bank. Yet it is common to speak of the bank note as though it were the right itself. When once he has been warned, the reader is not likely to have any difficulty over such usage, loose though it confessedly is.

Wealth and income. Fund and flow. There is a fundamental difference between wealth and income, which manifests itself in the matter of measurement. When wealth is measured, it is the quantity in existence at a certain time that is determined. In measuring income, we ascertain the amount of service that is rendered during a certain period of time. We do not speak of the yield of an orchard or the service of a carpenter at 1.25 P.M. of a certain date. It is ordinarily impossible to think of income except in reference to an appreciable length of time. An orchard yields so many bushels of apples per year, a mason lays so many bricks per day or per hour, etc. Wealth is a *fund*; it is measured as of a certain *instant* in time. Income is a *flow*; it is measured as of a certain *period* of time.¹

Value: A common denominator. Each kind of wealth, income, or property has its appropriate unit of measure. But there is one unit in which all kinds of wealth, income, and property may be measured and which is by far the most important of all units. This is the unit of value or price. The universality of exchange is one of the characteristics of the modern economic organization, to which much attention will be given in the succeeding pages. Value and price are, next to wealth itself, perhaps the most important of the fundamental concepts with which the student of economics must equip himself at the outset of his study. We must examine this subject closely and formulate our notions with clearness and precision, perhaps at the risk of introducing once more some fairly obvious concepts.

Transfer and exchange. When Mr. Jones gives his dwelling house to his wife, there is a change of ownership. The house, which formerly belonged to Mr. Jones, now belongs to Mrs. Jones. A *change of ownership of wealth* is called in economic terminology a *transfer*. Transfer takes place whenever wealth is bought or sold, bequeathed, or given away. Now suppose Mrs. Jones sells the house to the Brown Realty Co. for \$20,000. There has been another

¹ By the methods of the infinitesimal calculus, the period taken for the measurement of income may be reduced to zero, thus making possible the measurement of income as a *rate* of flow at an instant of time. This concept is of great importance in certain advanced problems of economics, as it is in some of the natural sciences. It need not further concern us in this elementary book.

transfer of the house. But this time there has been also a transfer of money, from the Brown Realty Co. to Mrs. Jones. These two transfers together make an *exchange*, which is defined in technical language as a *pair of voluntary transfers between two owners, when each transfer is made in consideration of the other*. A gift is not an exchange, since there is only one transfer. As people say, there is "no consideration."

Barter and money exchange. *Exchanges made without the use of money are called barter.* For example the farmer brings ten dozen eggs to the village storekeeper and takes in return ten gallons of gasoline and ten quarts of motor oil. This transaction fits the definition of an exchange. "Swapping" horses or anything else is barter. At some time in the history of any people barter was probably the usual form of exchange, such as there was, and a certain amount of barter always goes on among all peoples today. But the civilized world long ago abandoned barter as the normal form of exchange. In practically all exchanges today one of the transfers is a transfer of money, and money exchange has thus taken the place of barter. The subject of money will come up for special study in later chapters. For the present it is sufficient to know that *money is anything which is generally accepted in exchange for other things*.

Value. From exchange come two of the most important concepts of economics and business, not to say of modern human life in general; namely, value and price. As the term is employed in economics, *the value of anything is the quantity of any other thing that would be given in exchange for the first thing*. If ten dozen eggs could be exchanged for two bushels of potatoes, the value of the eggs is two bushels of potatoes, and the value of the potatoes is ten dozen eggs. If a ton of stove coal is sold for twelve dollars, the value of the coal is twelve dollars, and the value of twelve dollars is one ton of stove coal. It is not customary however to speak of the value of money in this way. On the other hand, values of practically all other things are regularly expressed in terms of money.

Price. *The price of anything is the amount of money that would be given in exchange for one unit of it.* If a pound of cotton would sell

for twenty cents, the price of cotton is twenty cents per pound. In stating the price of anything, the unit of measure used must always be stated or understood.

Value and price. The terms value and price are of course used more or less vaguely and inconsistently in popular speech, and there is some difference in usage among economists. The definitions laid down here agree as closely as is possible with popular usage, and they will prove themselves useful for economic analysis. As here defined, value and price are closely related to each other. They are distinguished from each other in that (1) value may be expressed in terms of any kind of wealth, property, or service, while the price of anything is expressed only in money, and (2) the term value is used for any quantity of a good, whereas price relates specifically to one unit. In practice value as well as price is usually expressed in terms of money; in which case value is price multiplied by quantity, and the value of one unit is identical with the price. Thus if a farm sells for \$25,000, we may say that the value of the farm is \$25,000, or that its price is \$25,000. Ordinarily we should not say that the price of fifteen tons of coal is \$150, but should say that the value of fifteen tons of coal is \$150, because the price is ten dollars a ton.

The unit of value. Value thus gives us a common unit in which to measure all kinds of wealth, income, and property. This is the unit of money value; *i.e.*, the value of the monetary unit, in America the dollar. Since value is almost always expressed in terms of money, we may ordinarily call the dollar the unit of value, not bothering always to say the unit of "money value." Thanks to this common unit, it is possible to total a list of various articles of wealth and property rights. Bushels of potatoes, yards of cloth, pounds of sugar, acres of land cannot be added together. But the values of all these things can be. It is possible therefore for the business man to make an "inventory," for the citizen to make up his "tax list," for anybody to make a list of his possessions and get the total. Above all, it is possible to keep accounts of wealth and property and income.

EXERCISES

1. Which of the following may properly be called wealth? Explain in each case.

- (a) A fountain pen. (b) An abandoned farm. (c) A factory building.
(d) A state highway. (e) The Gulf of Mexico. (f) A farm whose ownership is under dispute. (g) Fish in the ocean. (h) A moving picture actor. (i) A college dormitory. (j) An automobile on the junk heap.

2. Which of the following are wealth and which represent property rights? Explain in each case.

- (a) A share of stock. (b) A theatre building. (c) A bank deposit.
(d) A dollar bill. (e) A suit of clothes. (f) A life insurance policy.
(g) A delivery truck. (h) A football ticket. (i) An apple orchard.

3. During the last depression the United States Government has borrowed billions of dollars from the people and from banks through the sale of bonds.

- (a) Does this debt affect the wealth of the United States? (b) During the years these bonds are outstanding, interest will be paid the bondholders out of funds raised by taxation. Does the annual payment of interest on a national debt of this kind affect the wealth of the country? (c) Will the wealth of the country be affected if the bonds are paid off?

NOTE. This exercise refers to immediate effects, not to remote and indirect consequences.

4. In 1932-33 a number of states adopted laws providing for a moratorium on farm mortgage indebtedness. The holders of mortgages were unable to collect interest or undertake foreclosure.

- (a) Did such laws affect the wealth of these states? (b) Did such laws affect property rights?

5. What type of income is created by each of the following?

- (a) A farm. (b) A musician. (c) A watch factory. (d) A doctor. (e) A baseball player. (f) A fishing vessel. (g) A housewife.

II

THE FACTORS OF PRODUCTION

Production. The first step in the study of production must be to settle exactly what is meant by the term. Who are the producers? Shall all who engage in any form of mental or physical labor be considered producers, or is there a dividing line; are some engaged in productive and others in unproductive labor?

Common usage would lead us to say at once that the farmer, who raises wheat, is a producer; he obviously produces wealth, in that he creates something which did not previously exist. But a little reflection will show that he is a creator or producer only in the sense that he has been instrumental in effecting a change in the form or relationship of various physical elements already in existence so as to make them suitable for human consumption. The cotton manufacturer likewise adds nothing to the world's store of materials, but when he twists cotton fibre into thread and weaves thread into cloth he has added to the utility of the cotton in the eyes of the consumer. If we examine any other form of agriculture or manufacturing, we shall find that the essence of the particular operation is to bring about a change in the form of the good which will bring it one stage nearer the condition where it may satisfy a human want. Agriculture and manufacture, carpentry and blacksmithing, and innumerable other lines of activity, by thus changing the form of material goods, are active in the *creation of utility*.

But this is not the only way that utility is created. There may seem to be little similarity between farming and the operation of a railroad if one views only the technique of each business, but the purpose is essentially the same. Wheat on the farms of the West is quite as useless to the consumer of the East as if it had never been grown, and the service of the railroad in bringing it to the consumer is just as essential as that of the farmer in growing it. The grocer who keeps flour in stock to meet his customers' demands is performing a similar service, for the consumer not only wants a particular

good at a particular place, but he wants it at a definite time. Today's flour is patently more useful in satisfying today's hunger than flour which will not be available until next week. Likewise a productive service is rendered by those who facilitate the sale of goods. When A desires to sell wheat, he gives his broker a warehouse receipt for wheat stored in some elevator; the broker sells the wheat to B by delivering to him the warehouse receipt. A service has been rendered, for which the broker will collect commissions from A and B. The wheat is presumably one step nearer to the direct satisfaction of a human want. Utility has been created, by a change in ownership, even though the actual wheat has not been touched.

Thus are illustrated the four forms which the production of wealth may take: the creation of *form utility*, of *place utility*, of *time utility*, and of *ownership utility*. All effort directed toward the production of wealth may be classified under one or another of these four headings; agriculture and manufacturing result primarily in the creation of form utility; all the varied transportation services are engaged in creating place utility; merchants of all classes and bankers, brokers, and other middlemen are creating time utility and ownership utility. To the first one alone the name production is sometimes given, doubtless because of the obvious character of the change, but in view of the essential nature of the other three forms it is difficult to justify such usage, and we shall use the term production of wealth to include all four.

Although our primary interest in the study of production centres around the production of wealth, it should not be forgotten that man has wants which cannot be satisfied by wealth alone. A mechanical piano may satisfy the longing of one for music, but to another it is a sorry substitute for the performance of the pianist. If the labor of the pianist in making the roll for the mechanical piano is productive, certainly his performance on the concert platform must also be productive. We must therefore include in our definition of production the rendering of direct personal services by free persons. Completely stated, *production consists of (a) the production of wealth, which is defined as the creation of utility in wealth, and (b) the rendering by free persons of direct personal services.*

This definition is intentionally broad, including any effort which results in the satisfaction of a human want. In defining utility, it was pointed out that scientific definitions have nothing to do with questions of ethics or morals. Like utility, the word production carries with it no moral significance, although economists are not therefore unaware that effort spent in some directions may be less commendable than effort spent in others and that the results of human exertion may be fatal to individuals or disastrous to society. The study of production is the study of human activity in satisfying wants, whether or not that activity can be approved on moral or ethical grounds.

The factors of production. In spite of the fact that man, physically and intellectually, is today probably little different from what he was at the time when historical records begin, and in spite of the fact that the natural forces and laws have remained the same, the effectiveness of man's efforts devoted to the satisfaction of his wants has during the period covered by history immeasurably increased. The explanation lies in the fact that man has gradually accumulated knowledge of the natural environment and learned to make it serve him in production.

This power to guide and control natural forces to his own ends marks one great difference between man and the beast. Animals take what is offered them, and if their necessities are not supplied some or all of them must die. Man struggles so to change the environment that there shall be enough of the necessities to keep him alive; he forces nature to produce the plants and animals he wants. He has also learned how to make tools and machines which make his own labor more effective. Centuries before the beginning of historical times man must have roamed the fields seeking wild berries and edible roots for his food, and his early attempts to produce food were crude in the extreme. But with centuries of striving improvements have come, so that at present the same effort results in many fold the product of former times. The economic history of mankind is, from one point of view, the story of how men have secured food, shelter, clothing, and the comforts and refinements of life in increasing abundance and with less and less effort.

Man has enlisted in his service two powerful partners, and these

three — man, nature, and capital — work together in production, with results which, when compared with man's first efforts, are almost unbelievable. These three partners are commonly called the *factors of production* under the titles, *land*, *labor*, and *capital*, the first two being figuratively used to stand respectively for nature (not merely land in the narrow sense) and man. The rest of this chapter will be devoted to the study of their place in the economic organization.

Land. Land is essential in providing "standing room" for the population of the earth; it provides the food supply for the people and the raw materials for production and is the original source of all that man has. It is more than mere land, for it embraces the whole natural environment, including the oceans, lakes, and rivers, the topography of the land, whether a level or a mountainous country, the character of the soil, the amount and seasons of the rainfall, the range of temperature at different seasons, the mineral deposits, the water power, and a host of other features, all of which are important and may have a determining effect on the character of occupation and the type of industry in a particular region.

There are many examples illustrating the influence of the character of land on the industrial life of a locality. The conjunction of coal and iron in close proximity to each other almost inevitably stimulates the building of metallurgic works and leads to the growth of manufacturing; the point where one large river empties into another invites the building of a trading centre; the even humidity of the atmosphere may give a region an advantage in the making of fine cotton cloth; the presence of waterfalls in the past invited the founding of gristmills or cotton mills and at present determines the location of hydro-electric plants. While many industries are located in places which seem at the present time to offer few natural advantages, investigation of the cause of the original establishment will usually disclose that there was at one time some real advantage. At the time when cotton mills were first started in the United States, the only available power was furnished by running streams. Hence the mills were located in places where water power was available, such as Lowell and Lawrence. Today water power is not of such importance as steam power, yet we find the textile industry of

Massachusetts still largely concentrated in the region of its birth. The reason is that when an industry is once located in a place it attracts other manufacturers in the same line of industry. There may no longer be natural advantages in the locality, but there are artificial advantages, such as a labor force skilled in the particular trade, the presence of subsidiary industries which can take care of minor operations, transportation, banking, and other facilities devoted to that particular trade.

Labor. Land without labor is useless for the purposes of man. Even in the spots most favored by nature some effort is required to secure the goods desired, and on most of the earth's surface much exertion is necessary to make a living. For the majority of men continuous or regular labor is painful and is avoided if possible. Only when the desirability of things which can be secured by labor is considered to outweigh the unpleasantness of mental or physical exertion will the effort be undertaken. Mental effort is as truly labor as physical exertion. Probably the physician derives more pleasure from his work than a factory helper does, but our only standard of whether a given effort is labor or not is whether it contributes to production. The efforts of the physician, the lawyer, the judge, the business executive, and the government official all add to the total amount of enjoyable goods and services which constitute the income of society. Some of it directly increases the flow of goods and services, some indirectly by making the work of others more productive.

It has been suggested by students in this field that the efficiency of labor is dependent on three factors—race and inheritance, health and energy, training and social environment.¹ That the qualities which any race inherits are of great importance cannot be denied. A high degree of intelligence, the ability to learn, the ability to persevere are the recognized characteristic of certain races. Inability to receive more than a small amount of instruction or to grasp abstract ideas, fickleness of purpose, and a lack of the power of concentration are found in certain other races.

Health and energy seem largely to be governed by climatic conditions. Both a high average temperature and a low average tem-

¹ Cf. E. Huntington and F. E. Williams, *Business Geography*, p. 57.

perature impede mental and physical efficiency. Other things being equal a variable climate is more desirable than a monotonous one. A rise or a fall in temperature, or a storm, seems to stimulate activity. From this standpoint as well as from the standpoint of temperature the climate of the northeastern part of the United States is highly stimulating.

A good racial inheritance, health, and energy are of little avail without proper training. Such training is more important in the present era of the world's civilization than at any time in the past because of the greater complexity of the industrial and commercial processes. Even with the amount of training already provided there are doubtless thousands at work in positions for which they are not adequately trained and in which they are consequently not as effective as they should be. And perhaps more disturbing is the thought that many men and women are doing routine work who have the capacity to handle work of a more exacting nature more satisfactorily than some of those to whom it is now entrusted. The failure to make the most effective use of the talents of individual members of society constitutes a social loss; production is not as efficient as it might be.

The most reasonable hope of improvement in the use of labor lies in an expansion of the field of training and education, and particularly in the field of vocational education. To find out what a boy or girl is best fitted by natural endowment to do, whether medicine or engineering or blacksmithing, and then to provide the proper training is at the same time one of the most difficult and one of the most important of educational problems.

Capital. The third factor in production is *capital*. Ordinarily the term capital is synonymous with wealth — that is, useful material articles owned by man — but in discussing the factors of production it is convenient to limit the term capital to wealth which man has produced, thus excluding land. Land is distinguished from capital in this classification because of its peculiar relationship to production, but it should be remembered that it constitutes a particular and important class of capital in the ordinary use of that term.

At this point we are particularly interested in those capital goods

which are useful in production. There is an important group of capital articles which is desired primarily for its own sake rather than as instruments of production, including the mass of goods ready for consumption — food, clothing, furniture, etc. To this we give the name *consumers' capital* or *consumers' goods*, while we call the tools, implements, machines, and materials which are used only to produce consumable goods *intermediate goods* or *producers' capital*. It is however not always possible to draw a sharp line between producers' and consumers' capital, and the terms are not always mutually exclusive. Some goods are capable either of being used for further production or of yielding their satisfactions at once. Coal may be mined for use in factories, or it may be used in the kitchen stove with the double purpose of warming the room and cooking food, or it may be burned in the open grate. A ship may be capable of use as a yacht or as a carrier of freight. Moreover virtually every sort of consumers' capital may, when used to satisfy wants, be at the same time serving the ends of production. This is obviously true of the form of consumers' capital *par excellence*, food, and it will be recognized as true even of many of the less essential forms and even of many mere luxuries, such as athletic goods. But in spite of these qualifications we shall find the concepts useful in analysis, and with this warning the reader can easily avoid confusion.

Capital in production. We are at the present time so accustomed to tools and machinery of every description that it is difficult to imagine conditions in which they were entirely lacking; yet at some stage, long before the beginnings of history, man must have been dependent on the unaided efforts of his hands. Such a situation, where brute strength or cunning is relied upon to secure the means of subsistence, may be called *direct production*. What the first implement was is a matter of conjecture; it may have been nothing more highly developed than a club. Then possibly someone conceived the idea of attaching a sharp stone to the club, and an axe or hatchet was the result. Certain it is that tools came into use very slowly, and for centuries only the rudest and most primitive were known. Even during most of the centuries of which we have historical records, the machines and tools were of the simplest

sort. But with the first tool came a definite advance, for man had started on his yet unfinished process of changing the conditions of his environment to meet his rapidly growing needs.

Production with tools and implements is called *indirect* or *round-about* production. One may for example acquire the art of catching fish with his hands — this is direct production; but if he take time to make a fishing net he can greatly increase the effectiveness of his labor, including, as of course we must, the labor spent in making the net. Instead of proceeding directly to secure the desired products, time and energy are first spent in making tools, with the result that the production is made vastly more effective and, passing beyond the small range of activities open to one without tools, production brings forth things which could not conceivably have been made with the unaided hands.

We are not however particularly interested in the contrast between direct and indirect methods of production; the economies of the latter method are so patent as to require no further demonstration. The contrast of significance is between a less and a more intensive resort to the roundabout methods of production. The labor necessary for the erection of a modern flour mill may be spread over many years, the materials for its construction may be gathered from the four corners of the earth, and hundreds of men may be engaged on one or another of the preliminary processes, whereas the old-fashioned gristmill could be made in a few weeks' time by a relatively small number of laborers; but the output of the modern mill per unit of labor necessary to erect and operate it is far greater than that of its predecessor. Wherever we find an example of intense recourse to the roundabout or capitalistic method of production, we may safely conclude that it is because the effectiveness of labor is thereby enhanced. Capital instruments are labor-saving devices, not only from an individual's point of view, but from the point of view of society as well.

The accumulation of capital. Saving. The degree to which indirect methods of production may be carried depends, in part at least, on the quantity of capital which is available for productive purposes. Other elements enter, such as the progress of invention, the size of the plants in the industry, the demand for handmade

articles, and so on, but the quantity of capital is so fundamental a consideration that a study of the way in which it is accumulated will repay our consideration.

The formation of capital depends upon the readiness of men to postpone the gratification of wants in the present in the hope of increased satisfactions in the future; in other words, it rests on *saving*. In relatively simple cases this is quite clear. The savage who has gathered a supply of corn may be tempted to consume it before the spring, but if he can save enough for seed the following year he can hope for a more abundant supply in the autumn. A farmer today may use the proceeds from the sale of his crops to buy either a new automobile or an agricultural machine. The former will yield him pleasure at once, but the latter will enable him to expand his farming operations and to have a larger surplus of grain in the future. Nearly every manufacturer is compelled at some time or other in his career to choose whether to use the profits of the business to expand his scale of living or to "put them back in the business" in the form of enlarged facilities.

In these cases we can easily see that self-denial — saving — has added to the stock of capital. For the majority of people however the connection between saving and the productive processes is not so clear or so direct. Today the great mass of people gets its living by exchanging services for money and with this money purchasing the various articles which are required or desired. To the individual the saving of a part of his income does not seem directly to affect production; it appeals to him solely because of the necessity of providing against illness or old age or because money can be left in a savings bank or loaned or used to purchase stocks and bonds and can thereby earn interest. Yet it has precisely the same influence on production as saving by the manufacturer.

If we consider saving from the standpoint of society as a whole, it will be apparent that it entails the devotion to the production of intermediate goods of effort which might have been spent in producing goods capable of yielding their satisfactions at once. It further implies that the consumption needs of those working on the intermediate goods must be met from the flow of consumable goods produced by others. Were it not so they could not live.

We can imagine a simple kind of social organization where the tasks are arbitrarily divided among the different men and women, where all share in the common product, where the farmers are required to set aside a portion of their produce to support the miners and others working on intermediate goods, receiving their recompense eventually in the form of agricultural implements which will greatly increase the effectiveness of their efforts.

But although we may be able to find isolated cases where the intermediate worker is supported by direct payment in commodities, such as "grubstaking" a prospector, it is far from the normal case. The ordinary individual saves money, not commodities, and the laborer receives his remuneration in the form of money. This feature serves to obscure the nature of the problem and make it seem more difficult than it really is. A clerk who saves \$500 in the course of a year does so by denying himself the satisfactions which might have been obtained from the consumption of goods. When he places the sum in the savings bank, it is not kept in the bank's vaults until such time as he may require it but is loaned to some borrower. Now it is possible that the borrower may want the money to purchase goods for his own immediate pleasure, but most loans are made to men actively engaged in business and for business purposes. Let us suppose that the savings bank uses the money to buy a bond, a part of an issue of bonds put out by a railroad in order to be able to finance the construction of a branch line. The railroad uses the funds thus secured to engage men to work on the construction of the line and to purchase the results of past labor — steel rails, ties, etc. The clerk has actually passed on his control over consumable goods to the laborers engaged in the intermediate processes; by reducing his own consumption he has made it possible for them to work on non-consumable goods and yet to live. During the years which may be required to construct the railroad line the men engaged in all the preparatory stages might have been producing food, clothing, etc., and for the time being have greatly increased the flow of consumable goods, but the eventual result of abstinence on the part of some in society is the creation of a great machine of production, the railroad, which makes the efforts of workers in all lines many times more fruitful. The saver

gets his reward in the form of interest, but he also shares in the social gain, which consists of a more abundant flow of consumable goods.

It is difficult to overestimate the importance of the capitalistic or indirect methods of production and the necessity of saving, and it is easy to overestimate the adequacy of the volume of savings at present available in the world. Were the people of the nations of the earth suddenly to stop saving, spending all earnings on consumable goods, it would be but a short time before the entire structure of our civilization would be crumbling in the dust. On a smaller scale this happened in Russia. In the days before the World War Russia's industrial fabric depended in large measure on loans from foreign nations. Under the Revolutionary and early Soviet régimes Russia to a large extent was cut off from foreign sources of supply and in the disorder which prevailed was living in part on her capital. For a time there was practically no replacement of capital goods which had deteriorated or worn out. The result was an enormous decrease in the efficiency of Russian labor and the threatened extinction of many lines of industry.

The form of capital instruments. Invention. The effectiveness of labor in production depends not only upon the amount of capital available for use in production, but also upon the form which this capital takes. At the present stage in industrial development we could doubtless use to advantage a greater number of the tools and machines with which we are already familiar, but there must be some limit to the gain which comes from duplicating our present machinery. A more profitable line of development lies in improving the type of machines. Not two printing presses of the same capacity, but one improved press with twice the ordinary capacity.

The inventive genius of the world is constantly studying the possibilities of improving old machines and devising new machines and methods to take care of processes now performed by hand. Society profits by the inventions which make it possible to secure a greater flow of commodities with a smaller expenditure of effort. The saving which we have expressed in general terms appears to individual members of society in the guise of a lower price and therefore means that a given income represents command over a

larger number of commodities than formerly. Measures which stimulate inventive genius are generally in the public interest. The inventor may make a fortune in a few years through the monopoly which a patent right confers on him, but the people as a whole benefit still more. In the ordinary case the hope of making a fortune is the spur which leads the inventor to persevere through the period of fruitless experimentation. To those who understand, the wealth which Edison accumulated appears a small price to pay for the services he performed for society in placing so many comforts within the reach of people of modest means.

The capitalist class. Strictly speaking, anyone who owns capital is a capitalist, and there can be no clean-cut distinction between the popular terms "capitalist" and "laborer." Many a laboring man is the possessor of at least some small amount of capital, invested and yielding him an income in the form of interest or dividends. And the majority of the wealthy owners of great stores of capital are actively engaged in productive enterprise and so performing labor as that term has been defined. Yet vague though the distinction may be on the border line, it has become customary to speak of capitalists and laborers as separate social classes, and the classification and terminology have their purposes. In this sense the laboring class includes those who, though possibly owning some small portions of capital, nevertheless devote themselves to manual labor or the lower grades of mental labor and derive at least their principal support from the wages of their labor. The capitalist class includes those whose holdings of capital are relatively large and who are either above the necessity of performing labor or whose labor, if they do work, is of the higher grades devoted to the direction and control of capital and of the labor of others.

This class derives a considerable income in the form of interest and dividends from their capital, quite apart from any reward for their labor, and their position enables them to exert a powerful influence upon the destinies of the laboring class. To many the income and the power which result from the ownership of capital seem hard to justify. Justification or condemnation of the private ownership of capital is not our purpose. But it is pertinent to call attention to the essential service which, under that system, is

rendered by the capitalists and which the careless critic sometimes forgets. Capitalistic or roundabout production is a time-consuming process. Someone has to wait until the final product, the good which has led to all this preliminary investment in machinery and building, emerges. During this interval the workers must live. They are in reality producing a future good, but they must have present means of securing present goods. The capitalist is the one who does the waiting, and his operations resolve themselves into a series of advances to the laborers. The logic of the return which the capitalist gets in the form of interest is based on the fact that without this incentive he would not wait, would not tide the laborers over the preliminary stages of production, but would increase his own consumption. Savings would be checked, and society would lose as production fell off. The private receipt of interest could be escaped only by going over to some other social system in which there would not be private ownership of capital. Even so, saving would be required. Whether socialism or communism or syndicalism be attempted, the capitalistic method of production cannot be destroyed without relapsing into barbarism. Either individuals in society or the group as a whole must save and turn the savings into the machinery of production.

The entrepreneur: His function. Land, labor, and capital cannot work together without organization, direction, and control. Moreover in every productive undertaking there is *risk*; *i.e.*, the chance that the enterprise may prove not profitable and that the capital devoted to it may be lost and the human effort wasted. The owners of the several factors of production are, as such, not prepared to undertake this managerial function or to assume the risks of business. The landowner will provide land for a business if somebody will guarantee him a fixed rent. The laborer will work for an employer if the latter will by contract make himself legally liable to pay a definite wage. Finally the capitalist, insofar as he is performing the capitalistic function exclusively, lends money at interest to someone who can persuade him that the amount of the loan will be paid back in full. Someone must borrow the capital, decide what to produce, and employ labor and land to produce it, in the hope that the venture will be successful enough to pay him profits

over and above the fixed interest which he must pay to the owner of the capital. If the business is a failure, he must use his own resources to pay back the capitalist and to discharge his legal obligations to the landowner and the workers. *Anyone who performs this function of organization, management, and risk bearing is termed an entrepreneur.*

The entrepreneur is thus the man who makes the fundamental decisions as to what commodities to produce and how to combine labor, land, and capital in their production. He determines the degree to which it is profitable to supplant manual labor by labor-saving machinery, the amount of a given good which can profitably be produced. Whether a new invention is to be given a trial depends on his judgment. He studies the nature of the industrial operations, tries to simplify them as much as possible, and assigns laborers to one task or another according to his opinion of their abilities. The responsibility for the entire internal and external organization of a business falls on his shoulders — the coördination of production, finance, buying, and selling. The larger the business, the more incentive there is to delegate to others all tasks of a routine nature, and the more closely does the work of the chief executive partake of the nature of decision-making.

So important is this function in the modern economic organization that the entrepreneur is by some writers elevated to the rank of a fourth "factor of production." Since whatever the entrepreneur does falls clearly within our definition of labor, there seems no need of thus putting him in a class by himself. He is economically a laborer, although in the popular separation of social classes he generally falls in the capitalist rather than the laboring class. In fact the entrepreneur is quite generally a considerable owner of capital and so a capitalist as well as a laborer. We are able nevertheless to distinguish between the functions of laborer, capitalist, and business leader, even though they may be united in one person.

His reward. Business profits tend to vary directly with administrative and executive ability. Many business men never get more than wages — the wages which would be paid a hired manager to attend to the routine of the business; some do not even earn wages and sooner or later are forced out of the position of

business leadership; while others get business profits which correspond to their superior business ability.

There are those who regard the profits of the business man, particularly when they are large, with a disfavor no less than that with which they contemplate the interest received by the capitalist. While it is doubtless true in many instances that business profits reflect monopoly power and not a service to society, it is evident that much of the condemnation of business profits is the result of failure to appreciate the service which the business leader really performs. It is probable that as much native talent, not to say genius, is required of those who attain first rank in business leadership as is required of the pianist who becomes a performer of the highest order. There are many with modest musical gifts, but few with the natural ability, intellectual insight, and emotional qualities of Liszt or Rubinstein. We refuse to attend the public performances of indifferent performers, but we are compelled to permit many business concerns to be managed by men of mediocre business talent, because of the lack of a sufficient number of leaders of the highest grade.

It would seem therefore profitable to society to encourage in all ways the exercise of this business ability in order that the most capable leaders may serve as guides to the less capable. If the stage is ever reached where further progress is impossible, business profits will decline as the less capable of the business men catch up with the methods and devices of the more capable. As it is, the more talented business man stays several laps ahead and maintains the disparity between his income and the incomes of others by trying out new and possibly risky methods when his old methods of organization and production have become the common property of all. Thus is secured progress through developing more effective organization of the factors of production, from whose enhanced product the whole people stands to profit.

EXERCISES

1. The creation of a physical object may be unproductive; and there may be production when no physical object is created. Explain what is meant, and give illustrations.

2. Which of the following are producers? Explain in each case.

- (a) A stenographer. (b) A locomotive engineer. (c) A life insurance salesman. (d) A racketeer. (e) A dentist. (f) A mechanic. (g) A beggar. (h) A burglar. (i) A professional baseball player. (j) A fortune teller.

3. A, B, C, D, and E each saves \$1,000. A buys a piece of land. B deposits his savings in a bank. C invests his in the stock of a newly formed paper manufacturing enterprise. D spends his on a trip to Hawaii. E sets up an automobile repair shop. Explain how each of these activities affects the formation of capital.

4. Show how the factors of production are combined in each of the following enterprises:

- (a) A dairy farm. (b) A coal mine. (c) A retail store (d) A doctor's office. (e) A fishing enterprise.

III

INDUSTRIAL STAGES

Economic history. The purpose of this book is to give the student a knowledge of the essential features of our economic institutions and the principles which govern our economic life. The economic structure of today is so complex that without such an understanding one is not in position to pass judgment upon the present social order as a whole or to appraise critically proposed changes or reforms.

As a foundation for the more detailed analysis of modern economic society which will be found in later chapters it will be worth our while briefly to survey earlier systems which man evolved in his attempts to solve the economic problems of life. This will serve two purposes. Not only will it enable us to understand the modern organization better by knowing something of its antecedents, but it will give a background which will be helpful in evaluating modern institutions and methods. The reader will be enabled to grasp the fact that material progress has come slowly and painfully, that the degree of prosperity to which we have already attained is not lightly to be tossed aside or thoughtlessly jeopardized. He will be likely to acquire a healthy skepticism toward those who would bring forth the millennium at once by the passing of laws, while at the same time he will find abundant reason to anticipate future increase in material prosperity even more rapid than the past provided that the factors which make for progress are not stifled by artificial restraints.

The household system : General description. For our purpose we need go back only to that stage when the family or household was the economic unit, ignoring the struggles of primitive man which lie back of that period. The family, using the word in its broadest sense to include the clan or the tribe, lived a life unto itself. Most of the goods consumed by the family were made by the members of the group, for there was practically no trade with outsiders. There

must have been some division of tasks among the various individuals as natural aptitudes showed themselves, but the opportunity to develop special skills was very limited, and each individual was called upon to do a variety of tasks. Such a mode of life implies a limited number of commodities to meet the needs of life and a relatively low standard of living. Few families or households could be so situated as to command either the raw materials or the technical skill necessary for the fabrication of even a small fraction of the goods with which we are today familiar.

It was a simple type of organization indeed, lacking in many of the most common features of our modern life. There was virtually no exchange and hence little need for money; no wages were paid, as all were working for the common good and shared in the common product; the amount of capital used for production was limited to a few tools; the transportation involved was slight and gave rise to no problems such as vex the modern world. The organization was so foreign to our own experience that we have to turn to the remote past for an illustration. The large estates of the Romans might thus serve us, but a more interesting example is at hand in the manor, which was common in Europe in the Middle Ages and was the political and economic unit in England in the eleventh century.

The manor. The manor itself consisted of a village surrounded by arable land, pasture, waste, and woodland. Over each manor was a lord, who held it directly of the king or perhaps indirectly of some overlord on the basis of certain services, chiefly of a military character. Living in the village and cultivating the land which they held from the lord were tenants of various grades from the slave to the freeman.

The arable land was divided into two portions, that which the lord or his representative held, called the demesne, and the holdings of the tenants. The holding of a tenant might amount to thirty acres, while the holdings of the lord constituted a third or a half of the total arable land. Both the lord's and the tenants' holdings consisted of a number of acre and half-acre strips scattered over the tilled area, so that no one held his land in one block. The arable land was usually separated into three great open fields, one of which was planted to wheat or rye, another planted to barley, oats, or

beans, and the third allowed to lie fallow. The following year the field that had been allowed to lie fallow would be planted to wheat, the wheat field planted to barley, and the third field allowed to lie fallow. This was the rather crude method of conserving the fertility of the soil, made necessary by the lack of fertilizer and of knowledge of the proper rotation of crops.

One class of tenants was known as villeins. They had somewhat the status of serfs in that they were bound to the soil and were compelled to cultivate the demesne of the lord and to perform additional services for him at other times of the year, notably at the time of harvest. But the villein cannot be considered a slave, for he had his holding of thirty acres with a proportional share in the use of the waste, pasture, and woodland; he could not be sold from his holding, he enjoyed an independent family life, and he had the right to property which he might acquire by his own labor.

The manor was a world in itself, with very little relation to the world without. While not a family in the blood sense, it was a self-reliant household, with the lord at the head surrounded by a group of dependents. As can easily be imagined the most important occupation was agriculture, and it is no exaggeration to say that all of the laborers on the manor were agricultural laborers, although there was some slight development of crafts such as blacksmithing and milling. Practically all the food for the group was raised on the spot, the clothing was manufactured from the wool of sheep grown there, and the tools necessary for the cultivation of the soil and for such primitive arts as existed were of necessity the product of local effort. Trade in the modern sense scarcely existed, save in a few most necessary articles. Salt was always required, particularly to preserve meat through the winter, and in most cases it had to be procured from without. Tar was purchased to keep the murrain from the sheep. Animals also could be traded in since they provided their own means of transportation, but in general the difficulties of moving goods about were prohibitive. Even the kings and rich nobles who were the holders of several manors were compelled to come to the manors and consume the produce on the spot, since it could not be conveyed to them. The wage system was not to be found, nor can the payments which the tenants made to the lord be

considered the equivalent of modern rent, as they rested on custom and not on any real estimate of the worth of the land to the tenant.

Life on the manor may be made to seem attractive if sufficiently idealized, but as a matter of fact the modern worker would shrink with horror at the thought of being condemned to such a life. The semi-servile worker, who typified the largest portion of the population, could not leave the spot where he was born. His food was coarse and monotonous, his clothing the product of rude home industry, and his shelter squalid. Comforts, refinements, and pleasures simply did not exist for him. His wants were doubtless few, and fortunately so, for the means of gratifying wants outside those essential to life were entirely lacking. Such contentment as he found was not much above that of an unthinking animal which takes what it finds and looks no further.

The handicraft stage: The towns. With the rise in importance of the towns we enter upon a new stage of development, which has been called the handicraft stage. In the preceding period towns of importance were very few in number, but in the fourteenth and fifteenth centuries both on the Continent and in England they multiplied in number and in power. Many of them were founded under the protection of some powerful lord or ecclesiastical establishment; others seem to have grown up where trade routes met. In the early days of the town the population was primarily agricultural, the citizens cultivating their fields outside the town walls, and throughout medieval times the towns maintained in part their agricultural characteristics. But with the growth of the towns in population there came opportunity for certain men to develop special skill in particular crafts or trades, and we find artisans devoting the major part of their time to shoemaking, others to weaving, and so on. The products might be sold to townspeople, to people from the countryside, or even to traders from other towns. For the town as a whole the development of craft specialization meant the exchange of the surplus of its manufactured goods for the surplus produce of agricultural districts or of other towns.

Merchant guilds. That both the merchant and the artisan occupied places of greater significance than formerly can be seen in the formation, first by the merchants and later by the members of the

different trades, of associations or societies, called *gilds*. How far back the merchant gilds go is uncertain, but from the eleventh century on, as trade began to grow in volume and importance, there was a growth in the number and influence of these associations of merchants. Originally they were voluntary associations, entered into very likely for the purpose of mutual protection at a time when law and order had not developed to the point where it was safe for a merchant to travel about the country alone. Later on however we find them monopolizing the trade of the town and associating themselves with municipal authority in the attempt to maintain the monopoly of trade against outsiders. Indeed the relationship in many cases became so close that the merchant gild and the town authority coalesced.

Craft gilds. Shortly after the rise of the merchant gilds, craft gilds developed rapidly, and it was soon true that every craft or trade requiring any skill gave rise to a gild. In theory at least each gild regulated its own affairs in the public interest, and to ensure greater effectiveness in this a monopoly in each craft was granted, so that a man not a member of the gild might not ply his trade in the town. This requirement was not burdensome at first, for the gilds were interested in acquiring greater influence through a large membership, but later, as the position of the gilds became more firmly established, the monopoly power was used to protect the members against the possible competition of others, and admittance to the gild was made very difficult. Chief among its regulations in the interest of the community were those intended to secure honest workmanship. The right of gild officials to inspect the work and the workers' premises was recognized, and fines were levied for infringement of the gild regulations and for practices intended to deceive the purchaser as to the quality of the ware. Partially to secure youths thoroughly trained in the mysteries of the craft and partially to limit the number of men in the gild, a system of apprenticeship was instituted, so that a youth who desired to become a master weaver for example was required to serve with a master for a period ordinarily of seven years. At the end of that time he might be admitted to the gild and allowed to set up his own establishment, or he might continue to work for his master as a journeyman.

The craft gild presents a superficial resemblance to the modern trade union, but essentially it was quite different both in its constitution and in its aims. It is true that the aim of the gild was in part directed toward advancing the interest of its members and that the methods pursued closely resemble those to be found in some of the most highly organized trade unions of today, but as we have seen, the primary aim of the craft gild was to ensure honest workmanship and honest goods. Furthermore the trade union is composed of men working for an employer on machines and materials furnished by him and receiving wages, whereas the membership of the gild was made up of independent artisans, working at their own pleasure with their own tools and materials and securing their remuneration by the sale of their finished goods.

Medieval trade. The ordinary retail trade of the medieval town was in the hands of the traders or merchants, whose organization has been described. They maintained small shops or booths, to which came the citizens of the town to purchase all kinds of articles, farm produce from the country district, products of the town's industries, and "foreign" wares brought in from other towns or other countries. It must be remembered however that the craftsmen also generally had small shops or booths in the front of their houses, where they sold the products of their crafts to the townspeople. They also made goods "to order" for their customers. Manufacture and trade were thus not entirely separate, especially at the beginning.

The merchant class did not confine its activities to local retail trade. Besides the local shopkeepers there were traveling merchants. They performed the function of carrying the surplus product of the town's manufacturing industries to other towns and other countries for sale. And they brought back to their home town a great variety of foreign products purchased in the neighboring towns or in more distant parts of the country or even in foreign lands. The traveling merchant of the Middle Ages has his parallel in the modern peddler, whose horse-drawn wagon, filled within and bristling on the outside with a bewildering collection of pots and pans, brooms and knives, dress goods and toys, and countless other articles, was once a familiar visitor in every American village and

farmside and who even now may occasionally be encountered on the roads of the more backward country regions. He is a relic of the type of traveling merchant who in the Middle Ages furnished the one essential link between the products and the consumers of different towns and different nations.

The town market. The centre of the marketing organization of the Middle Ages was the town market, established in every town. Its principal function was to facilitate trade between the townspeople and the residents of the surrounding country. The market was a necessity because trade was still on so small a scale and so imperfectly organized. Today a farmer goes to town whenever he has produce to sell and finds it convenient to make the trip. He can always count on finding a wholesale merchant or broker to whom to sell his produce. He is also sure of finding retail stores always open and ready to sell him any ordinary thing that he may want in the way of local or imported products. In the medieval town there was not enough trading to call forth these facilities. It was therefore agreed between townspeople and country folk to set aside certain days upon which to meet at a certain place for the mutual exchange of their wares. Thus originated the "market," the "market-place," and the "market-day."

Markets very similar to those of the medieval town may be found today in many cities of Europe and the United States. For example certain streets or squares are often set aside for the produce market. Here gather each morning or on certain days the truck gardeners of the region surrounding the city, each with his wagon or motor truck loaded with vegetables and other produce. Here come the agents of the city grocery and provision stores, the commission brokers, the representatives of hotel and club stewards, and occasional thrifty housewives, to make their purchases. In some cities the market is inside a building, in which are rows of stalls, booths, and tables, for the sale of various products. Certain specialized markets will often be found, such as fish markets, flower markets, etc. In some countries of Europe and of South America the town market is of much greater importance than in the United States and bears a closer resemblance to the medieval institution. Often it is not continuous or daily, but is held only on certain days of the week, Sunday morn-

ing being a favorite time. As in the medieval towns, these markets serve as the meeting place for the exchange of town and country produce and are an important part of the economic life of the community. But while this sort of market may be a fairly common and sometimes significant institution in modern times, in the medieval town it was universal and extremely important.

The medieval fair. Another aid to trade in the medieval period was the "fair." The holding of a fair generally depended upon permission from the king in the form of a special charter. Such charters were granted to feudal lords or high church dignitaries. Fairs were usually held only at important places, especially at cities which were junction points in the national routes of trade and travel and at the important ecclesiastical centres. They were generally held in connection with some gathering, frequently a religious festival, which brought large crowds of people together from a more or less widely extended territory. The fair might last for any period from a few days up to six weeks or more. Certain of the great fairs, held regularly at stated dates each year, came to be national or even international institutions, attracting great numbers of important merchants from far distant places in addition to the local traders and populace. The fair was an occasion for social diversion and merrymaking, as well as for business. Besides the booths and stalls of the traders, there were side shows, wild animals, trained dogs, magicians, musicians, and "freaks," resembling closely the side show attractions clustered about the main tent of a modern American circus.

The duke or bishop under whose auspices the fair was conducted generally endeavored to make the fair popular and to promote its growth in every possible way. He frequently secured for the visiting merchants exemption from the customary tolls and taxes, immunity from attachment of the person for debt, and a general freedom from the prevailing restrictions upon trade. Hence the fairs often came to be the only places where there was any approach to real freedom of trade, and they were consequently known as "free fairs." The noble patron also was active in compelling fair dealing among the traders, enforcing contracts, and protecting purchasers from fraud, by a great variety of rules and regulations.

Special constables and even special sheriffs and courts were present to enforce the rules of the fair. Fairs in which the freedom of trade was particularly liberal and the regulations especially fair and well enforced gained a deserved reputation and flourished accordingly. It should be recognized that the motives of the noble patron were not entirely disinterested. Special fees and taxes were collected from the merchants who traded at the fair. The patron of the fair was often granted a monopoly privilege, through which trading elsewhere than within the fair grounds was forbidden during the period of the fair. The revenue of a large fair was frequently a handsome sum.

These great fairs thus became the regular meeting places of merchants and traders from near and far. They furnished a clearing house for all the wares that entered into the national and international commerce of Europe, such as silks, fine fabrics of wool and linen, leather, skins and furs, jewels and various manufactured articles, wines, spices, foreign foodstuffs, animals, and slaves. Here the merchants from the neighboring towns could, in return for their home town products, purchase foreign wares to take back and sell in the local markets.

Of course the fair is not unknown today. In certain of the less developed countries great fairs still exist, organized on the medieval model and performing a similar function. In the United States the fair, while still a trading device of some importance, has come to be more of an institution for exhibition and entertainment, though its advertising possibilities are fully appreciated by the manufacturers and merchants, and in this way it serves the ends of trade. The modern fair in most progressive countries is only a dim reminder of past greatness.

Municipal regulation of commerce. In this period the town rather than the manor was the economic unit. On some parts of the Continent the town tended to be the political unit as well as the economic unit, but in England, in theory at least, the authority of the king was superior to municipal authority. In practice the towns enjoyed a degree of local autonomy which is inconceivable to us. The authority of the town and the sympathy of its citizens were all on the side of increasing the wealth and prestige of the town. It was

a patriotic duty to protect it against the possible competition of the countryside and of other towns. Only if we remember this fact, can we understand the spirit animating municipal legislation in this age.

Among the most important of such regulations are those governing commerce. Commerce was conceived of as a privilege of a citizen of the town, to be extended to others only grudgingly and with reservations. Each town had its customs barriers, which prevailed against other towns and the countryside. A foreigner (who might be from the next town) might sell his goods at the market, but only under surveillance. He must sell only to townsmen, or to others after the citizens had satisfied their requirements; he must then have a citizen of the town act as intermediary, lest two foreigners conspire against the interests of the town, and he must expose all of his goods at once so that the total stock might be known. There were many other regulations of a more general nature to protect the interests of the townspeople. Selling goods on the way to market was forbidden; buying up the entire supply of a good to sell at a higher price and buying to sell again without adding anything to the value were likewise forbidden. The prices at which certain goods might be sold, the manner in which debts might be collected, the system of weights and measures, and a host of other matters directly or indirectly connected with the town trade were controlled by town authority.

The beginning of specialization. If we contrast this period with the one preceding, it is evident that there have been changes of considerable moment. In the first place, there is the beginning of specialization in production, of the differentiation of one occupation from another. Individually each is less self-reliant; he depends in part at least on the productivity of others and secures some of the articles he uses by making more of one article than he can use and exchanging his surplus for the goods of others. This development of exchange is significant, for exchange involves in some manner an evaluation of the two products, an agreement as to the relative worth of them, in other words value. The market, a place where buyers and sellers can meet to determine price by exchanging goods, was created. And a still more fundamental change can be seen in the growing necessity of money as a medium of exchange. Money

had been used previously for the payment of fines and for hoarding, but with regularly established habits of exchanging goods it became increasingly desirable. If one purchases few goods he can doubtless exchange wheat for shoes and cloth after a fashion, but if he depends on exchange to secure his whole stock of goods there must be some common unit in which the prices of all goods can be expressed.

Growth of freedom. It should not be thought that the development of town life was without its influence on life in the country. The citizens of the town were free as opposed to the semi-servile status of the rural inhabitants. Before the growth of the towns there had been little chance for an agricultural laborer to improve his condition, for there was no place where conditions were better. If he ran away to another manor, he would either be returned to his lord or might have to accept even worse conditions than those from which he had fled. Now he could go to a town, and as the custom held, if he remained there a year and a day he became a freeman. This raised the status of the agricultural laborer, as it made him more valuable in the eyes of his lord. Increasing use of money and familiarity with money payments, together with the demand from the towns for the surplus which the tenants might produce, brought about another change. This was the substitution of money payments for the personal service on the lord's demesne.

There were other factors and forces working in the direction of raising the status of the agricultural workmen. Into these we cannot enter, but we must note the result. In some Continental countries serfdom lasted until the eighteenth and nineteenth centuries; in England it had died out by the close of the sixteenth century. This is of significance, not merely on sentimental grounds, but because a free population is more energetic and ambitious than a servile one. England's development in the seventeenth, eighteenth, and nineteenth centuries was in large measure based on the superiority of a free population.

Judgment of the town economy. There can be no doubt that the period of town economy marks an advance over the previous stage. Certainly in the realm of material things there had been progress. The specialization of workers and the growth of trade brought within the reach of the country population many articles of com-

merce which they had not known before. The influence of the towns in raising the status of the agricultural laborer was decidedly progressive. The towns themselves offered means of amusement and enjoyment which had been lacking entirely or open to relatively few. On the important religious festivals the guilds united in giving "misterie" plays, so called from the name "misterie" or craft (from the French *métier*), to the delight of themselves and their audiences. The market was a source of profit not only in providing a place for the exchange of goods, but in bringing people together. There was therefore manifest an increased productive capacity and the possibility of satisfying a wider range of wants and, not least of all, a wider mental horizon.

It would be a mistake however to think of this epoch as the golden age, as some are wont to do. A medieval town was at all times picturesque, but probably at no time an attractive place of residence. The streets were excessively narrow, unpaved in many cases, and served not only as thoroughfares but as the dumping ground for refuse of all sorts. The houses were damp, cold, badly lighted and ventilated, and in many cases indescribably filthy. Plagues were a natural and inevitable concomitant of the unsanitary arrangements. Perhaps the independent craftsman got more pleasure from his work than the factory hand of today, but even that can be questioned. Certainly we know of nothing which would invite us to try to return to the handicraft system.

The domestic system. As time went on there was a development along two lines. In the first place, a greater degree of specialization by the individual in his particular trade manifested itself, and secondly some of the crafts were divided into distinct trades. This can be seen in weaving, where fulling and dyeing became separated from weaving. These changes were doubtless imperceptible at the time, but by the sixteenth and seventeenth centuries we see industry organized on a new basis.

Relatively little change is noticeable in the technique or methods of manufacturing. In weaving the workers still work in their homes at their hand looms with relatively few assistants. We see however that the craft guilds have lost some of their power over the trade and that weaving and other crafts have found roots in the countryside.

But the chief change has come in the control and direction of industry. The worker is no longer an independent craftsman working directly for the consumer but executes orders given him by a merchant, who supplies the materials on which he works and in some cases the looms as well. The craftsman has become a wage earner.

The explanation of this change is somewhat obscure in its details, but the general outline is clear. In the handicraft stage the artisan worked for a market which was at hand. The consumer himself gave him the order for the cloth or the boots. In the course of time the market for his wares expanded with the decreased isolation of the towns and with the increased importance of English foreign trade. Instead of production for a known and familiar market, production for an unknown and distant market was being substituted to an increasing degree. The craftsman was a maker of goods, not a merchant. He had neither the knowledge nor the ability to handle the marketing of his wares in distant parts, and there was therefore a definite call for the services of a middleman. The merchants stepped in and shouldered the risks of marketing.

Foreign trade was expanding, large profits were being made by the merchants, but the opportunities for investing these savings were rather limited — confined in the main to lines with which the merchants were personally allied. It was but a natural step from buying the products of a manufacturer to giving him orders, supplying the material, and even furnishing the instruments of production. Further profits could be made by stimulating production, and savings were available for this purpose. Some even went so far as to collect workers in separate buildings, thus anticipating the factory system, but in general there was only the modification of the handicraft system which has been described.

By the early part of the eighteenth century the domestic system was common in all lines of industry, although the independent small master survived in some lines, notably in certain of the woollen districts. This system lasted until it was supplanted by the factory system in the early part of the nineteenth century. Examples of it survived in the United States down to recent times, particularly in some phases of the garment industry. Vests and pants were given out, chiefly to newly arrived immigrants, to be stitched or button-

holed in the home at very low rates of pay. The exploitation of the immigrant in this way became at one time notorious. On the Continent in some of the specialized branches of the textile industry, such as lace making, it still persists.

This period is of peculiar significance, not only because of the breakdown of the customary methods of production, but because it is the beginning of capitalistic production in a new sense. We have already defined what is meant by capitalistic production, and of course production in the handicraft period was capitalistic. The difference here lies chiefly in the divorce of the capitalist and the worker. The master craftsman worked with his apprentices and journeymen. The merchant-manufacturer, as we may call the member of this new class, did none of the manufacturing himself, but he owned the materials and the finished goods and generally assumed the risks involved in production. It is the beginning of the separation of the worker from ownership in the tools of his craft, a step which is definitely taken in the factory period, and one which is responsible for some of the bitterness of modern labor problems.

The factory system. The factory system, which is characteristic of present-day industry, came to the fore in the last part of the eighteenth and the first part of the nineteenth centuries, though isolated examples of factories are known much earlier. Its advent is closely connected with that movement or development to which the name of "Industrial Revolution" has been attached. We shall study the Industrial Revolution in the following chapter and naturally shall have to consider the causes of the growth of the factory system and its chief characteristics. For the sake of completeness, however we may here briefly contrast it with the domestic system and point out its main features.

In the first place, there is the development of machinery. It must not be supposed that in the period of the domestic system there were no machines; the spinning wheel, which is as truly a machine as the most complicated modern automatic spinning machine, was known and used at a very early period in history. But now the simple machines worked by hand, which were all that existed in the previous stage, are replaced by complicated power-driven machinery. In the second place, the machines, instead of being in part at

least the property of the workers, belong to the factory. Thirdly in place of working in the home the workers are gathered together in a factory building. This implies a regularity of hours of work and a discipline during work which did not exist previously. And in the fourth place, the factory system brings with it a still greater separation of tasks into minute operations, many of which can be performed by machinery.

The factory system has given rise to many of the problems of modern economic life. The wage system, which is an inevitable part of it, the divorce of the worker from the direction of his work, the decline in value of the technical skill of the worker as the consequence of the increased use of machinery — all of these and other features have created problems for which no satisfactory solution has yet been found. Yet as we look over the period covered by our survey — more than a thousand years — we cannot escape the conclusion that progress has been steady, that man has improved his lot. Increased production both in quantity and in variety has given him a greater range of satisfactions; steady progress from a semi-servile status to that of a free citizen has been accompanied by an enlargement of the scope of his activities and a widening of his range of interests. Too narrow a view of the strife and struggle of conflicting forces of the day may tempt one to conclude that the world is on the downward path. A broader view gives greater hope for the future, for man seems to have that within him which keeps him continually striving for better things, material, intellectual, and spiritual.

IV

THE INDUSTRIAL REVOLUTION

It will be remembered from the previous discussion that the domestic system of manufacture was current in England during the first part of the eighteenth century. This system involved scattered workers, who used their own or hired machines in the cottages in which they lived and worked on materials supplied by a capitalistic manufacturer.

Nature of the Industrial Revolution. The eighteenth century witnessed changes in the technique of production so profound as to revolutionize industry. The period in which these changes came is called the period of the Industrial Revolution. This term is used, not to indicate a sudden change, but to denote the character of the change, for the transition, which started around the year 1770, had not completely altered the character of industry until about 1840 or even later. The Industrial Revolution involved the displacement of the domestic system of manufacture by the factory system, a great extension of division of labor, the invention and perfection of machinery in all lines of manufacturing, mining, and transportation, and consequently an enormous increase in the output of industry. It further involved social changes of great consequence, the congregation of workers in large factory towns, the subjection of labor to the discipline of the factory, the beginnings of labor organizations, the rise of a new class of capitalists, and in England a new political alignment. In almost every phase of social, political, and economic life conditions were altered, new relationships brought into being. Most of the problems of economic and social life of today find their origin in this period, and the student who would understand modern problems and try to evaluate modern institutions and relationships must include in his study a survey of conditions in England in the eighteenth century.

Why the Industrial Revolution began in England. Let us see first of all why the Industrial Revolution came in England and not in

some other country, why conditions in England made her particularly ripe for a change. In the first place, contrasting England with the countries of the Continent, there was personal freedom. Serfdom had disappeared in England by the end of the sixteenth century, while it persisted on the Continent through the eighteenth century. In the rural parts of the Continental countries it would have been difficult if not impossible to secure the necessary labor force for work in the factories, and the use of water power made location in the country necessary at times. The gild system still had control of industry in parts of the Continent, and the gilds bitterly resented any innovations which might threaten their supremacy, whereas in England their monopoly of the trades had pretty well broken down by the seventeenth century. This made it possible for the business man to embark on enterprises as he saw fit, without being controlled and disciplined by members of his own craft. The relative freedom of the English producer from government regulation and restriction, as compared with the Continental producer, favored experimentation with new modes and methods of production.

The spirit of freedom was also manifest in the abolition of customs barriers between towns and in the political and religious security of the individual. England had made not a little progress industrially by serving as a haven of refuge for skilled workers suffering from religious persecution on the Continent. Finally her position of geographical isolation rendered her free from fear of foreign aggression and contributed toward making her a favorable field for industrial changes.

There were still other factors of great importance. England had a large accumulation of capital, secured in the main from her foreign trade, which could be invested in manufacturing. Factory production with the use of expensive machinery is necessarily roundabout or capitalistic production. The foreign markets which the British merchants were serving were large and expanding; this meant that if more goods could be produced they could be sold. There were men trained in trade on a large scale, and production on a larger scale than formerly would not offer the difficulties that it might elsewhere. Finally there were vast deposits of iron and coal located conveniently to each other. Probably the absence of any one of

these factors would have retarded the development which took place, but coal and iron were without question essential, for they were fundamental to the application of power machinery to mining, transportation, and manufacturing.

Coal and iron. During the early part of the eighteenth century the iron mining industry of England was in a period of depression, largely because of the difficulty of securing charcoal for smelting the ore. In all England there were in 1737 but fifty-nine furnaces, which produced 17,350 tons of pig iron. In the same year about 20,000 tons were imported, chiefly from Sweden. A cheaper and more efficient fuel had to be discovered before any great expansion in the production of iron could be expected. Coal had been known for some time and was used to a slight extent for domestic purposes, but it had been unsuccessful as industrial fuel. The urgency of the demand for fuel led however to further experiments, and with the development of a method of coking the coal and the invention of the blast furnace, coal assumed an importance in industrial life which it has maintained to the present time.

The practical solution of the fuel question led to another problem, how to get the coal out of the mines. So long as little coal was mined, shafts near the surface could be used, but the sinking of deep shafts was impracticable because of flooding with water. This difficulty was partially solved by the invention of the steam engine early in the eighteenth century and the improved steam engine of Watt later on, which made it possible to pump out the water as well as to haul up the coal by power. The development of coal and iron production in England is shown in the table on the following page.

The figures for coal and iron are not given for the same years, but there is a fair basis of comparison. In the last half of the eighteenth century the production of iron increased seven-fold, that of coal more than doubled. In the first half of the nineteenth century iron production increased twenty-four-fold, that of coal somewhat over six-fold. These figures are interesting not only in showing the absolute increases in both iron and coal, but also in pointing out the increased efficiency in the use of coal. When we realize that coal was being used for power to an ever increasing degree, a twenty-

four-fold increase in iron production coupled with a six-fold increase in coal production points clearly to improved processes.

TONS OF COAL AND IRON PRODUCED IN ENGLAND¹

<i>Coal</i>		<i>Iron</i>	
YEAR	PRODUCT	YEAR	PRODUCT
1700	2,148,000		
1750	4,773,828	1740	17,350
1770	6,205,828		
1790	7,618,728	1788	68,300
1795	10,080,300	1796	125,079
1854	64,700,000	1854	3,100,000

The textiles: Woollen and cotton industries. England's greatness in foreign trade and her domestic prosperity in the sixteenth and seventeenth centuries have been ascribed to her woollen manufactures, and it was a cardinal feature of her commercial policy to protect her domestic wool supply, her manufactures of wool, and her foreign markets. A part of the woollen manufacture was domestic in origin, but it had been greatly stimulated by the influx of refugees from the Continent. The immigration of Flemish weavers who settled in and around Norwich in the sixteenth century had been encouraged and had helped develop that region into a prominent manufacturing centre.

The cotton trade in the first half of the eighteenth century was insignificant and was confined to Lancashire, later to become the seat of an enormous and profitable cotton industry. Most of the raw cotton came either from the Mediterranean and Oriental countries or from the West Indies, but the amount imported was small. Yet the first important inventions in textile manufacturing were applied to cotton first and then slowly to wool. The reasons for this are not difficult to understand. The cotton industry was new, and there was maladjustment between spinning and weaving. For the former the supply of laborers was insufficient, particularly with the growing demand for domestic cottons consequent on the limitations placed on the import of India cotton goods. There was a famine in cotton yarns; the weavers could easily use more yarn

¹ L. C. A. Knowles, *Industrial and Commercial Revolutions in Great Britain during the Nineteenth Century*, 1921, pp. 70-71.

than could be turned out by the primitive spinning methods. The development of cotton planting in the United States and the assurance which the invention in 1793 of the cotton gin gave of a continuous supply of raw material stimulated investment in cotton manufacturing. Furthermore the fact that it was a new industry gave the producers more freedom, for they did not have to face the hostility which would unquestionably have arisen in the case of the older and better established industry. It was probably easier to apply the new inventions and methods in the woollen industry after they had proved successful in the cotton industry.

Spinning. Spinning presented the greatest immediate problem. The old-fashioned spinning wheel which drew and twisted the fibres into a thread could make only one thread at a time, and it is said that six spinners were required to supply yarn for one weaver. It is therefore natural to find the first inventions in current use in this branch of the industry. In 1767 Hargreaves invented a hand machine, the spinning jenny, so named in honor of his wife, which enabled a spinner to make eleven threads at once instead of one and which was soon improved so as to be able to spin one hundred threads at once. This was followed the next year by Arkwright's water frame, a power machine as the name suggests, which spun a stout cotton thread that could be used for the warp (the thread which is placed lengthwise in the loom and which is crossed by the woof) in place of the linen thread previously used. And in 1775 Crompton perfected his mule, a hand machine, though soon adapted to power, which was a combination of the principles of the jenny and the water frame. This could spin thread fine enough for muslins.

Weaving. The increase in the output of yarn made possible by the improved spinning machinery and the increase in the supply of raw cotton reversed the situation which had existed earlier. Instead of a famine in yarns there was a surplus great enough to admit of a considerable volume of exports — a thing considered highly undesirable as it tended to stimulate foreign weaving. Attempts were made to attract weavers from the woollen trade, but attention was also focused on improving the weaving machinery. Weaving offers technical difficulties. As has been explained, the woof is passed

through the warp, and to do this the warp must be capable of adjustment so that the woof passes over one thread and under the next; the succeeding thread of the woof will reverse the process, thus binding the whole together. In 1733 Kay had invented his flying shuttle, which simplified the task of passing the woof through the warp and permitted a man to do his work in half the time and enabled one man instead of two to weave the widest cloth. It met however with opposition and was not introduced until the latter part of the eighteenth century. Cartwright had invented a power loom in 1784, but it was imperfect, and it was not until the first part of the nineteenth century that a really successful power loom was devised.

The use of power and the growth of factories. Factories with power machinery were not an immediate result of these inventions. The first of the new devices were hand machines and were used in cottages, as had been the case with the earlier hand machines. It is stated that spinning continued to be carried on in the homes on the jennies twenty to thirty years after they were invented. But as the spinning machinery was improved and became heavier, power was substituted; first of all horses were used, then water power, and then steam.

The use of power almost inevitably meant the factory system, with machinery, power, and materials supplied by the owner, and the laborers under a common discipline. In the first place, the machinery was becoming too expensive for the workers to purchase; it had to be supplied by one with large capital resources. Again to use power economically it had to be applied to all machines at the same time, and this involved restraint on individual freedom, the fixing of common hours and conditions of work.

The factory system however spread very slowly. The new machines did not always work well, and when water was the source of power it offered serious drawbacks. There might be an insufficient supply, an oversupply, or no supply at all. It is said that by 1790 there were only 150 water frames in use, so great was the disinclination to start factories.

The improvements in the steam engine effected by Watt in 1776 and 1782 offered an improved source of power, and steam gradually

supplanted water. But the steam engine was by no means the steam engine of today; it was imperfectly built, broke down frequently, and was anything but economical in the use of fuel. In 1800 there were steam engines in only fifty-three manufacturing centres, and as late as 1835 there were 1,297 water wheels to 1,953 steam engines.

It has been stated that woollen manufacturing did not apply the new machinery until after it had been used in cotton manufacturing. Several reasons for this have already been suggested, but there were other reasons as well. Wool offered greater technical difficulties than cotton, in that its fibre is more difficult to handle. Woollen fibres are softer, less uniform, and less tractable than cotton fibres. There are great differences in wools, and every fibre has individual peculiarities to a greater extent than is true of cotton. There was also not the scarcity of hands which was manifest in the cotton trade. Throughout the centuries in the growth of this trade a fairly accurate adjustment had been made between the raw material and the workers necessary to make it up. The increase in the supply of raw material because of the imports of Australian wool did not come until after 1830.

Nevertheless the use of improved machinery and the development of the factory system were inevitable. The jenny for spinning woollen yarn came into use about 1785, and power spinning in factories in worsteds had started by the late years of the eighteenth century, although it did not develop in wool spinning until early in the nineteenth century.

In speaking of the development of the factory system in the woollen and worsted industry one writer says: ". . . the adoption of factory organization and the introduction of machinery came very slowly. There were scarcely twenty factories in Yorkshire in 1800; the power-loom was not introduced into Bradford till 1826, when it was the cause of fierce strife and riots; combing was done by hand until well into the forties, and many technical difficulties rendered it undesirable to use the power-loom in the woollen industry until about 1850. Writers in the middle of the last century speak of the wide-spread existence of the cottage system, and the memories of people still alive reach back to the days when the hand-

loom was to be found in almost every cottage. Thus we come to the conclusion that the Industrial Revolution had little more than its beginnings in the eighteenth century. The great change came first in the cotton industry, then in the manufacture of worsteds, and lastly in the making of woollen cloths. In the Yorkshire branches of the textile industry, the revolution did not actually take place until the nineteenth century; the face of Yorkshire had been little altered by 1800, and a half a century had still to elapse before it could be claimed that the factory and power-driven machinery had displaced the old hand methods."¹

It must not therefore be thought that the growth of factories meant the entire elimination of the home spinner or weaver; they persisted well into the nineteenth century. As late as 1833 a Parliamentary committee estimated that there were 200,000 hand-loom weavers of cotton in the United Kingdom. In the beginning the factories would seem to have supplied the increase in the demand, leaving most of the weavers unmolested. Eventually of course the superior economies of factory production made it impossible for a hand weaver to get a living, but there were many side lines in which hand weaving persisted to the second half of the nineteenth century. The stocking industry for example had not been transformed into a factory industry, and in the forties it was doubted if it ever would be, so great was the difficulty of applying power machinery to the making of stockings, and so firmly attached to their present mode of living and so opposed to going into factories were the workers.

Reluctance to establish factories. The slowness in the establishment of factories may seem strange to us of the present day. The economies of the factory system of production seem so obvious that it is difficult to appreciate the real opposition which existed a century ago. Let us remember, in the first place, that we are accustomed to change, that we welcome changes which have a fair chance of proving profitable. The worker of the eighteenth century had no such point of view. The customary method of doing a thing was the right method, and violent changes were to be avoided. Furthermore the worker was fairly well satisfied with the situation as it existed. He might have a small plot of ground with his cottage and

¹ H. Heaton, *Yorkshire Woollen and Worsted Industries*, 1920, p. 283.

do a little farming on the side; he possibly had the right to pasture his cow in the common pasture. His wife and children worked with him and were self-supporting, and the family wage was fairly large and stable. He was his own master, could work when and how he chose, and found the discipline of a factory very distasteful. It is small wonder then that the factory was looked upon with disfavor as a disrupting force and that riots to destroy the factory machinery broke out in several places.

The opposition to factories appeared however not only on the part of the workers. The employers were none too eager to try them out. The "putter-out" of goods was really in a strategic position. He incurred no loss if he stopped giving out orders. His capital outlay was small compared to that necessary in a factory, unless he owned the machines. In that case he secured a rental from the worker and could at times use his position as a giver of orders to force the worker to pay a high rental. He was only gradually forced to turn to the factory system, either by a shortage of labor in certain lines or because of dishonesty in the use of the material which he supplied or difficulty in securing the goods when he wanted them. In some cases strikes on the part of the domestic workers caused him to establish a factory. But when the process was once under way the success of the great leaders forced others, employers and employees alike, into line.

The development of the factory system, granted the inventions and improvements necessary to make it possible, was not because of the enthusiasm of the age; it was rather the result of economic factors which were well-nigh irresistible in their force. And of these the chief factors can safely be said to be the increasing foreign demand for goods and the necessity of supplying these markets in order to provide means for continuing the continental wars.

Machines and tools. We may appear thus far to have treated the Industrial Revolution as if it were confined to the textile industries. True the development of the textile industries and particularly of the cotton industry offers an impressive example, but many other industries as well were affected. Indeed the development of the textile industries depended in a very vital way on the developments in other industries. Throughout the whole period there was con-

stant action and interaction. Advances in one line caused advances in another line; these advances made possible still further progress in the first.

The early hand machines, the looms and the spinning wheels, were made of wood, and the first of the new spinning machines were also made of wood. Among the numerous disadvantages of wood are the size which is necessary to secure strength, the impossibility of permanent precision in moving parts, and the difficulty of securing the requisite strength whatever the size. The latter difficulty was especially prominent when water and later steam power began to be used. Wooden machines could not stand the strain, and therefore first the movable parts and later all the parts of the machines were made of iron.

But the fabrication of iron machines, either for the textile industry or in the form of steam engines, was a new industry. There were few trained workers, the tools for making machinery were inadequate, and it is not surprising that parts made by hand did not fit when put together or that continual breakdowns were the rule. It is said that when a new machine was sent to a factory the foreman of the shop which built it might have to spend a month fitting the parts together, easing up here and filing off there, until all the parts of the machine were adjusted to each other.

It was essential that machine tools should be made before any great development in the building of machines could be anticipated, for without machine tools precision in cutting and boring — making parts which are interchangeable — is impossible. The development along this line did not come until the first part of the nineteenth century. In part the demand for engines and machinery stimulated this development, in part was stimulated by it.

Transportation : Roads. The growth of the engineering industry and the increase in the output of goods of all sorts put a strain on the transportation system in two ways. There was, in the first place, the question of transporting the finished goods and the raw materials and, in the second place, the hauling of coal from the mines and to the centres of manufacture.

The roads of England in the early eighteenth century were in very bad condition, and the naturally slow method of transport by

wagon was rendered still slower. There were, it is true, waterways on which goods could be shipped, but they were not sufficiently developed nor well enough linked up with each other to suffice for the growing needs of commerce and industry.

The first step in the development of transportation was improvement of the roads. In the eighteenth century turnpike trusts were created which gave to private persons the right to charge tolls for the privilege of passing over the particular stretch of road. They had the corresponding duty of building the road and keeping it in repair. The roads which formed the main highways were kept in fair shape, though naturally there was a great deal of difference between roads, and in winter most roads fell into a shocking state of disrepair. At the opening of the nineteenth century road-making received a great impetus from the improvements of McAdam, who showed how a durable road surface could be made, and from Telford, who was skilled in road engineering. In addition the turnpike trusts tended to form amalgamations and to reorganize so as to operate more efficiently. During the first part of the nineteenth century there was steady progress in road-making and repair until the advent of the railway plunged many of these trusts into bankruptcy.

Canals. While these improved roads were fairly satisfactory as a means of transportation for manufactured goods which had a high value, their use was prohibitively expensive for bulky raw materials. At this time there was a growing demand for coal for industrial and household use, the potteries required clay, the factories in Lancashire needed a cheap method of transporting raw cotton. This need was met by the building of a network of canals, which, together with the rivers, joined the coal-producing areas, the industrial areas, and the sea. The first canal was built in 1761 by the Duke of Bridgewater to connect his coal fields with Manchester and was soon followed by another canal connecting Manchester and Liverpool. Then came a rapid succession of canals, built by private persons who collected tolls from the users. By 1830 there had been built 1,927 miles of canals and 1,312 miles of improved riverways, and there were also 812 miles of open rivers in England and Wales. A barge of twenty tons could navigate from one end of England to

another. The result was a cheaper and more rapid method of transportation. Without the canals the development of industry would probably have been considerably delayed.

Railways. The railway period, which really opened in 1826, had been preceded by experimentation in rail wagon carriers using either gravity or horse power. The rails were constructed of wood until 1767, when iron rails began to be substituted and proved to be much superior. There had also been experimentation with stationary engines to haul cars. These early railways were built to carry coal from the coal mines to the canals or in some cases to towns. The Stockton & Darlington Railway was chartered in 1821 and was the first railway to use steam locomotives and to carry passengers, but the Liverpool & Manchester Railway, which was chartered in 1826, inaugurated the steam railway period. In 1825 Stephenson's "Rocket," though not the first locomotive, was the first to demonstrate clearly the practicability of the locomotive.

The development of the railway transportation system in Great Britain is a subject into which we cannot enter, although it forms a very interesting and important chapter in railroad history. It is worth noting however that the railway was introduced and extended in England under great difficulties and in the face of much opposition. Some of the current objections to the steam railroads are given in the following paragraph, and although these arguments today seem fanciful, it should be realized that they had great weight at the time and served to increase the difficulties which the promoters of railways had in securing charters from Parliament.

"The country gentleman was told that the smoke would kill the birds as they passed over the locomotive. The public was informed that the weight of the engine would prevent its moving; and the manufacturer was told that the sparks from its chimney would burn his goods. The passenger was frightened by the assertion that life and limb would be endangered. Elderly gentlemen were tortured with the notion that they would be run over. Ladies were alarmed at the thought that their horses would take fright. Foxes and pheasants were to cease in the neighborhood of a railway. The race of horses was to be extinguished. Farmers were possessed with the idea that oats and hay would be no marketable produce; cattle

would start and throw their riders; cows even, it was said, would cease to yield their milk in the neighborhood of one of these infernal machines.

“Vegetation, it was prophesied, would cease wherever the locomotive passed. The value of land would be lowered by it; the market gardener would be ruined by it. The canal could carry goods cheaper. Steam would vanish before storm and frost; property would be deteriorated near a station. It was called the greatest draught upon human credulity ever heard of. It was erroneous, impracticable, and unjust. It was a great and scandalous attack on private property, upon public grounds — one class was informed that the locomotive would travel so fast that life and limb would be endangered, another was told that it would be too heavy to travel at all.”¹

The development of the railway transportation system not only made possible progress in other lines by providing a means for shipping raw materials and finished goods cheaply and expeditiously, but it also stimulated the development of the coal and metallurgical industries, adding as it did to the demand for both coal and iron.

Summary of effects of the Industrial Revolution: Industrial progress. The Industrial Revolution gave rise to new industries and expanded previously undeveloped industries. The cotton industry, mining, the iron-working trades, and transportation are examples of this. It caused the break-up of the domestic system of manufacture and its replacement by the factory system with power-driven machinery. New districts in England became important industrial centres. There was a tendency thus for industry to concentrate in the north of England where the iron and coal fields were located. Towns of a somewhat different type, with large industrial populations, sprang up, and from this crowding together of workers in the factories and in the homes developed a spirit of class consciousness greater and more powerful than ever before. The laborer was further removed from his employer than formerly; this too tended to weaken old ties and added to the growing class consciousness. The substitution of machines deprived the laborers of a part

¹ L. C. A. Knowles, *Industrial and Commercial Revolutions*, 1921, p. 256. Quoted.

of their competitive strength, for it made trade skill and knowledge of less value.

Along other lines great changes were taking place. A new group of capitalist manufacturers was being formed, whose interests lay in an extension of manufacturing and freedom of trade. They wanted as little regulation as possible and cheap food for their manufacturing population. Economic interest and the philosophy of the age worked in the same direction; *i.e.*, toward abolition or lack of enforcement of laws which might hinder the progress of manufacturing. A striking example of the victory of this new class over the old vested interests is seen in the abolition of the tariffs on the importation of grain in the first half of the nineteenth century.

England's productive powers showed a rapid and consistent increase. In almost every line increased production can be noted, and in some instances the increase was truly phenomenal. A table showing the growth in the output of coal and iron has already been given. In 1801 the export of pig and bar iron amounted to 4,584 tons; by 1850 it had grown to 611,407 tons. In 1780 the official value of cotton goods exported was given as £355,060; by 1800 it had increased to nearly £6,000,000, and by 1850 to over £110,000,000. In consequence of a duty imposed on printed cottons we know the quantity printed until the duty was repealed in 1831.

<i>Year</i>	<i>Number of yards printed</i> ¹
1796	20,600,000
1800	32,870,000
1814	124,600,000
1830	347,450,000

Foreign commerce, both exports and imports, showed an almost steady increase. In fact by the middle of the nineteenth century England was in a position of unchallenged supremacy in production and foreign trade.

Paralleling this growth in productive power and in some respects a necessary condition of it, we find a growth in consuming power and a widening of the market. England's manufactures were sent to all corners of the globe and among her best customers was the

¹G. R. Porter, *Progress of the Nation*, ed. by F. W. Hirst, 1921, p. 305.

United States. The United States was increasing in population and in wealth; there were more people to buy goods, and they were better able to pay for them. The development in transportation tended to bring this and other markets nearer to the producers, as the cost of shipping goods to and from the ports decreased and as ocean-going steamers began to supply more regular and cheaper service than the sailing vessels.

Producers who had in the past been supplying a fairly steady market were dazzled at the prospect before them, and there was a race between producers to see which could turn out the largest volume of goods and reach the market first. They were impelled to this course by the knowledge, which they soon acquired, that they could reduce the unit cost of the good which they were making by increasing the quantity produced.

We should not neglect to note that changes in the methods of marketing and financing manufactured goods were also consequences of the Industrial Revolution. In the face of continual increases in the output of the factories, old methods which had developed to handle a relatively small and stable quantity were found to be inadequate and gave way to new methods designed to meet the new situation. Discussion of this topic however must be postponed until later chapters.

Effect on the laborers. We should however err were we to leave the impression that the period covered by the Industrial Revolution was one of universal prosperity and contentment. Such was far from being the case, and in many respects the period forms one of the least attractive in economic history.

The hand workers were in a fairly comfortable and contented condition in the first half of the eighteenth century, but by its close conditions had begun to change for the worse. The growth of factory towns made it impossible for the worker to have his plot of ground and do a bit of farming on the side; in fact the workers were so crowded together that the living conditions were miserable. Houses were located in the midst of filth and disease, with practically no attention to sanitation and with little attention to the water supply. The factories were damp, hot, ill-lighted, and ill-ventilated, and adequate guards for the machinery were practically

unheard of. The competition for employment enabled the employers to exact twelve, fourteen, or fifteen hours of work a day. Women and children were exploited mercilessly; it was not unknown for children of seven to be forced to work at a machine for twelve hours a day, and women were compelled to work as long hours as men at tasks far beyond their strength.

Perhaps a lowering in the standard of living of the workers was inevitable; a period of transition is always difficult, for there is both ignorance of the real conditions and a helplessness in the new situation. We can consider the situation which developed as a temporary incident of a great economic change, requiring time for old customs, traditions, and ideals to adjust themselves to new conditions. The prevailing philosophy taught that unrestricted competition was desirable in the public and private interest and doubtless made the readjustment less rapid than would otherwise have been the case. It is well recognized today that public policy requires the state, for its own safety and health, to preserve the health and vitality of the children, but it only gradually dawned on the people of that generation that regulation of the hours and conditions of labor of women and children was necessary.

It should also be noted that the children of the hand workers had led no life of ease before the factory was open to them. Almost as soon as they were able to walk, they were set to work by their parents and were expected to earn their living. There was little opportunity for pleasure and none for education in a formal sense. There was almost as much reason to protect the children from the tyranny of their parents as from the exactions and ill treatment of factory foremen. Any factory inspector or employment manager at the present time knows how ready many parents are to lie about the ages of the children, how ready they are to sacrifice their health and education for the addition which the children can make to the family income.

The limitation of the number of hours which children could work and the prohibition of factory work before a certain age were forced as soon as the public was fully informed of the conditions, and this created a child unemployment problem, the solution to which was found in state education. In the long run therefore the factory

system worked for the benefit of children, but not before it had taken its toll in death and disablement.

For the men too the factory system probably meant an advance, once the period of transition was over. The factory became a better place in which to work than the cottage had been. Factory inspection insured cleanliness, air, and light. The work was monotonous, it is true, but scarcely more so than working at a hand loom, and with a shortening of the hours of labor the discipline of the factory, though always distasteful, became less burdensome. Work was likely to be more regular than in the domestic system, for the factory owner dared not run the risk of dispersing his labor force during a dull period, and because of his heavy capital investment he could not afford to allow machinery to lie idle if it were possible by any means to keep operating. As time went on workers found it easier to combine into effective unions and force improvement of wages and working conditions. This meant the beginning of labor troubles, but few would deny that the improvement in the life conditions of the mass of workers has been worth while.

The Industrial Revolution has meant not only a rise in the standard of living of the workers but, notably in England, has permitted an enormous increase in population. Formerly England was of necessity largely self-sufficient, but the development of her manufacturing industries in conjunction with the opening up of virgin soil in the New World made possible the exchange of manufactured goods for food supplies and raw materials and thus enabled the population to expand far beyond the limits which would be set by the natural productive powers of the soil. The population of England and Wales, which amounted to about eight and one-half millions in 1801, had doubled by the middle of the century and had increased to almost thirty-eight millions by 1921.

For society as a whole the conditions of life have been improved. The struggle for the bare necessities is not so bitter, and the range of comforts and luxuries attainable by the masses of the people is enormously wider. Less engrossed in the business of making a living than were our medieval predecessors, we are more free for living.

The Industrial Revolution in America. It is questionable whether it is proper to speak of an industrial revolution in America. Cer-

tainly the improvements in manufacturing which were altering the industrial and social life of England in such a radical fashion had no immediate counterpart in America. The reasons for this are easy to understand. In the first place, manufacturing, save in the home and for home use, was hardly established in the colonial period. The people could more easily secure their manufactured goods by devoting their time to developing their natural resources and exchanging the products of the field and forest for the goods they required. In the second place, the colonial policy of England did not contemplate the development of manufacturing. It was in fact definitely and consistently suppressed. The colonies were conceived to be a source of demand for English manufactures, and not a rival to them. England also tried to maintain a monopoly of her improved machinery by stringent prohibitions on exportation. Hence at the beginning of American national history we find a very great dependence on foreign importation.

The first great stimulus to the development of manufacturing came when the flood of foreign goods was shut off by the embargo of 1807 and later by the interruption to trade caused by the War of 1812. A cotton spinning mill had been established in Pawtucket as early as 1789 by Slater, who had been an apprentice under a partner of Arkwright and built his machines from memory. But by 1805 only three more had been started. After the embargo other spinning mills were founded, and in 1814 Lowell founded at Waltham, Massachusetts, the first factory using power looms. Following the war there was a period in which the newly established industries were favored by relatively high customs duties, and the manufacture of cotton, woollen, and iron products grew in importance.

The United States did not however offer a favorable field for the rapid spread of the factory system. The richness of her natural resources was such that it was unprofitable to spend more than a small portion of the available capital and labor in developing manufactures. A practical illustration of this is seen in the difficulty of securing labor for factory work. In England it soon came to be a choice between working in a factory or having no work at all; but in the United States up to fairly recent years there was always the

opportunity to secure land in the West at a nominal cost. As a factory laborer a man could expect little more than a modest living; as an independent farmer he might count on a fair living and could reasonably hope for financial independence for himself or his children. Similarly the development of the nation's mineral resources and of transportation diverted capital and labor from manufacturing.

In New England and other regions not richly endowed by nature, manufacturing increased in importance with the growth of population and capital. Yet if we look for the period in which manufacturing ranks with agriculture in importance, we shall find it in the second half of the nineteenth century and not in the first half. We see the results of the Industrial Revolution in America, and America has utilized and improved the inventions of the period, but the growth of manufacturing here has been a gradual process, unaccompanied by the industrial upheaval manifested in England.

V

THE FORM OF THE BUSINESS UNIT

Control of production in the modern world is largely in the hands of the entrepreneurs, who organize and direct industry, induced to do so by the expectation of gain in the form of business profits. In the course of the extraordinary era of economic progress ushered in by the Industrial Revolution, old ways of conducting business have been modified, and new forms of business organization have been introduced in order that the several branches of industry might adapt themselves to changing conditions and might function more easily and efficiently. The form of the business unit and its manner of functioning are matters of considerable significance to society.

The individual proprietorship. A business which is simple in nature, which requires little capital, and to which attaches relatively little risk does not require a complex form of organization. In fact complexity of organization would prove more a handicap than a help. The small retail grocer or baker for example generally finds the individual proprietorship best suited to his needs. In this form of organization the proprietor is in sole charge of the business, responsible alike for its success or its failure. He may start up in business when he chooses and cease operations when he so desires, without the bother and expense of legal formalities. Unless an activity is specifically prohibited by law, such as the coining of money or the carrying of mail, no line of business is closed to him. Responsible to himself alone, he is in position to make decisions at once without seeking either the advice or the assent of an associate, and he is thus able to take advantage of the opportunity of the moment. His business secrets cannot become common property through the indiscretion of associates.

While there are advantages for the small business in this form of organization, there are certain drawbacks which make it unde-

sirable for a concern of any great size. In the first place, the individual is seldom able to command a large amount of capital, for he is rarely able to invest as much in his own business as can be secured by a partnership or a corporation. And if he is able to do so, the great risk involved is a deterrent, for he is personally liable for all debts of his business. Hence at the present time we find relatively few large enterprises organized and operated by a single individual. While the number of manufacturing plants owned and operated by individuals is over one half of the total number, the output of these plants forms an insignificant part of the total output. In agriculture alone the individual is still the predominant figure.

The partnership. The partnership is a more complex type of organization than the individual proprietorship, for it involves relations between the individual partners as well as the relationship of the partnership to the public. Essentially it is a voluntary association of two or more individuals for the performance of a specific or a general object. This object may be accomplished in a day or a week, or the association may be formed for more general purposes, such as the establishment and operation of a factory, and may last a longer time.

In any case the purposes for which the partnership is formed, its duration, the ways in which it may be dissolved, and the mutual rights and obligations of each partner to the other will be the subject of an agreement, usually set forth in written form. As was true in the case of the individual proprietor, the partnership can be created with little legal formality and without delay.

While the partnership agreement controls the relations between the individual partners and can determine the share of profits going to each partner and the proportion of the common losses which each must pay, it cannot control the relations with the general public. From a legal standpoint each partner is considered to be an agent of the partnership, fully empowered to bind the partnership in all lawful undertakings. In the event of inability to meet all debts from the funds of the partnership, each partner is liable to the full extent of his personal resources for the debts of the business, even though the agreement may make one partner

responsible for only a certain proportion. For example let us assume that the partnership agreement of Jones and Brown stipulates that Jones shall receive seventy-five per cent of the profits and be responsible for seventy-five per cent of the losses. Large losses occur and the debts exceed the resources of the partnership. Jones finds himself unable to meet his share of the losses, and therefore Brown must make up the difference if he is able to do so. This is known as *unlimited liability* and is one of the chief points of difference between the partnership and the corporation.

The fact that there are several men united together makes it possible for the partnership to control a larger amount of capital than can the individual. In fact many partnerships have been formed when the individual had reached the point in the development of his business where additional capital was necessary and could most easily be secured through the admittance of an associate. Few concerns are able to finance themselves with their own capital exclusively; they must rely to some extent on bank credit, secured through loans from banks, and on commercial credit, obtained by purchasing goods to be paid for in the future. The ability of the partnership to secure capital in these ways is likely to be greater than that of the individual, in so far as the unlimited liability of several men offers a greater margin of safety to the creditor than the unlimited liability of one.

Another advantage of the partnership over the individual proprietorship may come from the association of abilities. An important question will be viewed from more than one point of view, and the decision which is reached may therefore be sounder. Offsetting this there is the possibility of serious disagreement and consequent delay in making decisions, but at least snap judgments are likely to be avoided. A certain degree of specialization is possible, as in a wholesale house where one partner is entrusted with the buying, another with the selling, and the third with financial questions. This feature is especially prominent in partnerships of professional men, such as those formed by several doctors, each of whom may be a specialist in a particular branch of medical practice.

One of the chief weaknesses of the partnership is the prospect of its enforced dissolution on the death of a partner. Under the

common law doctrine,¹ unless modified by statute law, the death of one of the partners automatically dissolves the partnership and makes necessary an accounting and distribution to his heirs of the deceased partner's share of the net resources of the partnership. This may be a hardship, both to the partnership and to the heirs. But the heirs cannot join the partnership without the formation of a new partnership.

The theory on which this doctrine is based is that the partnership is a peculiarly personal relationship and that anything which occurs to interfere with this must necessarily bring the original arrangement to an end. A logical corollary is that a partner may not transfer his interest in the partnership without causing a dissolution of the partnership, because no associate can be forced on the other partners against their will.

But the partnership has such obvious advantages over the individual proprietorship that, before the corporate form was devised and made common, it was the prevailing form of organization in enterprises of moderate and large size, and it is still not uncommon to find large concerns conducted for many years under a partnership agreement, although the corporation has steadily increased in favor and now overshadows the partnership in importance.

The limited partnership. A special type of the partnership is known as the *limited partnership*. This involves one or more general partners, who manage and direct the business and have unlimited personal liability, and one or more limited partners, whose liability is limited to the extent of their investment in the business. While the right to associate in business as partners without special sanction is well recognized in our law, the formation of a limited partnership is a privilege conferred by statute law, and certain formalities are prescribed, such as registration and the giving of public notice of the limited liability of certain partners. It should be noted furthermore that limited liability does not imply freedom from liability; it means rather that in case of losses the resources of the limited partner are not all available for the settlement of the

¹ Common law is based on universal custom which has received recognition and sanction in decisions rendered by the courts. Statute law consists of laws enacted by a legislative body, such as the Congress of the United States or the legislature of one of the states, and enrolled on the statute books.

debts of the partnership. Usually the liability is limited to the investment in the business.

Another characteristic of the limited partnership is that those partners with limited liability must take no share in the control of the business. If their relation to the business is such as to lead the public to believe that they are actively engaged in the business, the plea of limitation of liability may not be accepted by the court in settling the affairs of the partnership.

This form of organization may prove more advantageous than the general partnership or even the corporation. It offers a method whereby capital may be secured without relinquishing control of the business. A partner anxious to withdraw from active business may do so, leaving his share in the business intact. A partnership which has been broken up by the death of a partner may be re-formed as a limited partnership, without disturbing the assets or breaking the continuity of operations, by admitting the heirs as limited partners. In states where the expenses of incorporation are heavy, where state supervision of corporations is quite strict, or where corporations are heavily taxed, the limited partnership may be a more desirable form.

The corporation : Modern growth. Though now the most significant type of business unit in the United States, the corporation is of comparatively recent development, the great increase in the number of corporations having come since the Civil War. In the fields of banking and insurance, mining and quarrying, railroading and other public utilities the corporation is practically supreme today. Even in manufacturing, where its growth has been slower, it has a position of dominance. In 1919, although only 32 per cent of all manufacturing establishments were owned by corporations, they turned out almost 88 per cent of the total value of the products. By 1929 the number of manufacturing establishments owned by corporations had increased to 48 per cent of the total. These corporate establishments employed 90 per cent of all wage earners in manufacturing, and produced 92 per cent of the value of the product. In some lines of manufacturing their importance is even more striking. For example in the manufacture of rubber boots and shoes corporations have entire control. Corporations turned

out 97 per cent of all the agricultural implements made in the United States and 99 per cent of the clocks and watches. In occupations formerly thought the province of the small retailer, the corporation has made its appearance, as is evidenced by the rapid growth of chain stores in the grocery, drug, tobacco, restaurant, and other fields.

Despite the continued growth in the number and importance of corporations it must not be thought that they have entirely supplanted other forms of business organization. In certain lines of industry such as agriculture, where the corporation is rarely found, and the building industry, the partnership or the individual proprietorship is still either the dominant or an important form of organization. Taking our economic activity as a whole corporations in 1929 produced 57 per cent of the total national income and other forms of business organization the remaining 43 per cent.¹

Nature of the corporation. A corporation has been defined in part as "an artificial being, invisible, intangible, and existing only in contemplation of law." This definition indicates one of the most distinctive characteristics of the corporation; in the eyes of the law it is a legal or artificial being, a fictitious person, conceived to have an existence apart from that of its owners. The partnership exists in the persons of the partners. For example suits by or against the partnership are brought in the names of the partners. But the corporation sues and is sued in the corporate name.

Furthermore the corporation depends for its existence upon the sanction of the law. The individual business man and the partnership carry on business by virtue of rights long recognized by common law, but the corporation exists only through the official act of the state. In the United States a special act of the state legislature was formerly necessary for the formation of each corporation, but there are at present general corporation laws which prescribe the method of incorporation.² If the organizers fulfill the requirements as laid down in the corporation laws and make application in due form to the proper state official, a *charter*, or *certificate*

¹ *Big Business: Its Growth and Its Place*. Twentieth Century Fund, Inc. New York, 1937, p. 17.

² Some corporations, such as the national banks, are chartered under federal laws.

of incorporation, will be issued to them creating the corporation and defining its rights and duties.

Called into being by the act of the state, the corporation exercises only such powers as have been conferred upon it. A partnership may engage in any activity which is not prohibited, but the corporation is restricted to the range of activities sanctioned by the terms of its charter or the general laws, either explicitly or implicitly, as essential to its existence. Acts beyond the scope of the corporation's powers are termed *ultra vires* and may result in the revocation of the charter and the dissolution of the corporation.

The doctrine of personal relationship, which is stressed in partnerships, is not so important in the case of corporations; the association of capital is of greater significance than the association of individuals, and this finds its exemplification in the right of each of the owners of the corporation to transfer his property rights in the corporation at will, without seeking the assent of the other owners, either individually or collectively. It follows that the death or disability of one of the owners of the corporation does not bring about its dissolution. This makes the corporation a more permanent form of organization.

In some cases indeed a corporation has perpetual existence, but it is customary at the present time to limit the life of a corporation to a definite period, such as twenty or fifty years. At the expiration of this period the corporation is automatically dissolved, but renewal of the charter is so easy to obtain that there is virtually perpetual existence. Dissolution may also result, as we have indicated above, from *ultra vires* acts of the corporation. As a matter of fact however the authority granted by the charter is usually couched in such general and comprehensive terms that the ordinary corporation is in little danger of exceeding its powers. Finally dissolution may be brought about by the action of the majority of the owners with the assent of the state authorities.

Another distinctive feature of the corporation, as compared with the partnership, is the *limited liability* of the owners of the corporation for corporate debts. This is not a necessary mark of the corporation; there have been corporations the owners of which

have had full personal liability for corporate debts, but it is usual at the present time to limit the liability of the owner to his investment in the business. A man who has purchased 500 shares of stock in a corporation is liable to lose what he paid for them, but the law does not lay hold of his personal resources to satisfy the creditors of the corporation, except in certain special cases where the law provides for double liability, as it formerly did in the case of stock issued by national banks. In this instance the owner of a \$100 share of stock of a national bank was subject not only to the loss of the sum he had paid for the stock but might also be called upon to contribute \$100 more to meet the debts of the corporation in the event of insolvency.

In concluding this brief survey of the characteristics of the corporation, we should note that the conduct of the affairs of the corporation is delegated by the owners to hired managers. The owners elect representatives from among themselves, called *directors*, to care for their interests and to supervise in a general way the operation of the business, but the active management of the corporation falls upon the officers, who are elected usually by the directors. Problems of importance are referred by the officers to the directors. Questions of policy or of unusual importance may be submitted to the owners collectively for action, although such reference is actually infrequent.

Capital stock. The capital stock represents ordinarily the original and permanent investment of the owners in the corporation. Suppose that a group of men incorporate the Steel Manufacturing Corporation and agree to contribute \$1,000,000, divided into 10,000 shares of \$100 each. A takes, let us say, 5,000 shares and pays \$500,000 into the treasury for them; B takes 1,000 shares, and so on until all the shares are sold. The corporation now owns \$1,000,000 in cash, and the organizers of the corporation have certain property rights against the corporation. The collective name for these property rights is *capital stock*, and if we say that the capital stock of the company is a million dollars, it is an elliptical way of saying that the property rights which the owners of the corporation hold against the corporation (on account of their original investment) amount to a million dollars.

Each owner of a share of stock, or *shareholder*, becomes a part owner in the corporation and as such is possessed of certain rights against the corporation, such as the right to receive dividends when legally declared, the right to share in the resources of the company if it is liquidated, and usually the right to vote at the meetings of the shareholders. A *certificate of stock*, signed by the proper officials of the corporation, serves both as evidence of the shareholder's rights of ownership and, when properly endorsed, as a convenient method of transferring his rights to another. It must be emphasized that a stock certificate is not synonymous with a share of stock; a share is one of the units into which the total property rights of the owners have been divided, and the certificate is merely presumptive evidence that a part of the total rights is vested in the holder. This is an example of the distinction between a property right and the document in evidence thereof which was pointed out in Chapter I.

Many corporations find that their financial needs are best met by the issue of two types of capital stock, which differ from one another in some important respects. These are *common stock* and *preferred stock*. Common stock usually carries with it the exclusive right of voting at shareholders' meetings. Preferred stock also may confer the right of voting, but usually the preferred stockholder has no voice in the management so long as his dividends are regularly paid. The ownership therefore of more than one-half of the outstanding common stock will give control of the corporation, since the owner has one vote for each share of stock he holds. In fact in a large corporation a much smaller percentage will usually suffice because of the lack of interest on the part of a number of the small stockholders and of division among the larger ones.

Dividends are declared on common stock by action of the board of directors when profits have been earned and when it is judged wise to distribute them, and the amount or the rate of the dividend is entirely within the discretion of the board. Hence the stockholders in a new corporation are uncertain of a return on their investment for some time, conservative policy dictating that the earnings be reinvested in the business, but as the corporation becomes more and more successful the dividend rate may rise until

it is much higher than the rate of yield on any other type of security.

Preferred stock differs from common stock in that a fixed rate of return is promised, and in view of this the rate of return is somewhat lower than that which can be expected eventually from common stock. Furthermore dividends must be paid on preferred stock before any can be declared on common stock. Ordinarily the preferred stockholder has no redress if the company is unable to pay dividends in any given year, and if during the succeeding year very large profits are made, he does not then receive dividends for the previous year. But there is frequently *cumulative* preferred stock, upon which back dividends must be paid up before any dividends can be declared on the common stock. This protects the preferred stockholder against manipulation of the accounts to conceal profits and avoid thereby the paying of dividends. Preferred stock may also be *participating*. This means that when a dividend has been paid to the common stockholders at the same rate as to the preferred stockholders, both the common and the preferred stockholders share in proportions already determined in any further dividend payments made for that dividend period.

The owner of this type of stock has preference over the common stockholder in another direction also; *i.e.*, in the distribution of assets in case of liquidation. The common stockholders receive their share of the assets only after the preferred stockholders have been paid in full. If the remaining assets are insufficient to redeem the common stock in full, the amount is divided pro rata among the common shareholders; if there is more than enough to redeem the common stock, then the remainder is usually divided pro rata among the common stockholders.

Bonds and notes. A part of the funds for the operation of the corporation are often secured through the sale of its bonds and notes. A bond is a certificate of indebtedness issued by a corporation, bearing interest at a stipulated rate and payable at stated times, and having a certain number of years to run before the corporation is obliged to redeem it. Essentially it differs from a promissory note only in the length of time before maturity, the type of security on which it rests, and the greater formality with which

it is issued. The notes of a corporation, apart from the promissory notes of a few months' tenor, partake of most of the characteristics of the bond. The majority of them however run for a few years only, whereas bonds may not mature for as many as forty years after their issue. The notes frequently pay a higher rate of interest than do the long-time bonds. This arises from the fact that note issues are resorted to at a time when the ruling rate of interest is too high to make it desirable to issue bonds; by the time the notes mature it is expected that the interest rate will have fallen so that, if necessary, long-time bonds may be sold to replace them.

The bondholder has no voice in the management of the business so long as his interest payments are met with regularity, except in so far as his contract with the corporation may limit its freedom of action in respect of matters which have a bearing on its ability to pay the principal or interest of the bond. If the security behind the bonds is a mortgage on the physical equipment of the concern, as is frequently true, default in the payment of either interest or principal may force the bondholders to foreclose the mortgage and take over the operation of the business to protect their own interests. In case of the liquidation of a business the bonds are redeemed before either preferred or common stockholders receive anything.

Stocks and bonds compared. The most essential distinction between bonds and stocks is that the former represent a definite contract on the part of the corporation to pay money to the bondholders, whereas stocks only entitle their holders to receive their shares of such business profits as may be earned and their shares of the corporation's net assets upon liquidation. As regards the risk of failure to receive payment of income and return of the amount invested, the bonds of any given corporation are obviously safer than its stocks. On the other hand, they carry no likelihood of very large returns (beyond the fixed interest rate). Finally they give no control over the conduct of the corporation's business. Preferred stocks carry more risk than bonds but give the chance of a somewhat larger return; they generally give no control. The common stockholders take the greatest risk but enjoy the possibility of the highest return; and they have usually complete control. Of course this does not mean that bonds are always safer than stocks. The

stocks of a sound, well-managed corporation would carry less risk of loss than the bonds of a corporation whose assets were of questionable value or whose management was inefficient or dishonest.

There is however another risk to which all investors are subject; *i.e.*, the risk of a decline in the purchasing power of the monetary unit. Here the bondholder bears the greatest risk. Both his income and his principal are defined in the contract with the corporation in terms of dollars; if the general level of prices should rise, the bondholder's command of commodities would decline. The shareholder, either by an increase in the dividend rate or through an appreciation in the market value of his securities, may be left in a financial position substantially unaltered or even improved. By the same token, the bondholder stands to gain from a fall in the general level of prices. The stockholder gambles more than the bondholder on the success of the enterprise. The bondholder gambles more than the stockholder on the price level.

The property account of a corporation: Principal items of assets and liabilities. Many of the features of fundamental importance in connection with the corporation can be illustrated to best advantage by the use of a simple property account, or balance sheet. We may define the property account as a record, as of a particular time, of a real or fictitious person's *assets* and *liabilities* and of the relationship between them. Assets consist of wealth and property rights against other free persons and their wealth; a person's liability is a property right of another against him or his wealth. Thus if A owns a house, it appears on the asset side of his account; his promissory note to B for \$500 appears on the liability side of his account and on the asset side of B's account.

Referring to our former example we will assume that the Steel Manufacturing Company has been incorporated with a capital stock of \$1,000,000, all of which has been paid for in cash. The property account would then appear as follows:

STATEMENT OF STEEL MANUFACTURING CORPORATION
JULY 1, 1929

<i>Assets</i>		<i>Liabilities</i>	
Cash	\$1,000,000	Capital Stock	\$1,000,000

Capital stock appears as a liability since the corporation is considered as a distinct entity, and its capital stock represents rights which the owners have against this artificial person. The cash appears as an asset, since the corporation owns it.

Now let us assume that the corporation markets \$500,000 worth of its bonds, and spends \$1,200,000 for plant, including machinery, etc. Its balance sheet would then read :

STATEMENT OF STEEL MANUFACTURING CORPORATION
JULY 31, 1929

<i>Assets</i>		<i>Liabilities</i>	
Plant	\$1,200,000	Bonds	\$ 500,000
Cash	300,000	Capital Stock	1,000,000
	<u>\$1,500,000</u>		<u>\$1,500,000</u>

We have now two types of liabilities — the capital stock, which is a property right of the owners, and the bonds, which are property rights of outsiders.

During the next ten years the plant is in operation making and selling the goods for which it was built. It has sold some of its goods and has purchased some of its raw materials on credit; it has on hand a supply of materials of various kinds, raw and finished, and it has increased its investment in the plant. Furthermore the years have been profitable, and we find this reflected in the balance sheet of July 1, 1939.

STATEMENT OF STEEL MANUFACTURING CORPORATION
JULY 1, 1939

<i>Assets</i>		<i>Liabilities</i>	
Plant	\$1,800,000	Bonds	\$ 500,000
Raw Materials	100,000	Interest due Bondholders	30,000
Finished Goods	150,000	Notes Payable	50,000
Accounts Receivable	75,000	Accounts Payable	60,000
Notes Receivable	25,000	Capital Stock	1,000,000
Cash	102,560	Surplus	500,000
		Undivided Profits	112,560
	<u>\$2,252,560</u>		<u>\$2,252,560</u>

Many of the items on this balance sheet need explanation. The items "accounts receivable" and "notes receivable" among the

assets are property rights which the corporation has against other persons and which are evidenced respectively by entries on the books of the corporation and by promissory notes. Presumably they arose from the sale of goods. The corresponding items among the liabilities refer to similar obligations to others on the part of the corporation.

The other two items which need explanation are surplus and undivided profits. It will be noted that since the date of the last statement, our assets have grown from \$1,500,000 to \$2,252,560, an increase of \$752,560. At the same time claims of outsiders against the corporation have grown from \$500,000 to a total of \$640,000 (bonds, interest due, notes payable, and accounts payable), an increase of \$140,000. The increase in total assets exceeds the increase in liabilities to outsiders by \$612,560. This represents a gain from all the corporation's business, and this gain belongs to the stockholders. The corporation is a fictitious person and as such can never have assets greater than its liabilities; they must always balance. To indicate the growth in the value of the stockowners' holdings we might possibly add the \$612,560 to the capital stock, but it is desirable to leave this item intact, in part to show the amount of the original investment, and partly for technical reasons. These two new items are therefore introduced to represent the growth in the owners' share in the assets of the corporation.

The only distinction between surplus and undivided profits is that which is dictated by convenience. Surplus is kept at a round figure and usually represents a reinvestment in the business; undivided profits are not definitely committed to any use but are generally considered as available for the payment of dividends, and their amount varies continually as profits are earned and paid out or losses suffered.

It should be emphasized that surplus and undivided profits are not cash or money or any other form of wealth. They are property rights. The sum of capital stock, surplus, and undivided profits represents the accountant's estimate of the total worth of the property rights of the owners (the stockholders) against the corporation and is known as the *owners' equity* in the corporation, or the *net*

worth of the corporation to the owners.¹ The net worth may always be found by subtracting the liabilities to all others than the owners from the total of the assets. The reader will observe that, in any property account or balance sheet, the asset side is a list of property rights against wealth and persons, whereas the liability side shows to whom these property rights as a whole belong. The liability side is thus a record of *propriatorship*.

Losses and insolvency. The relationship between these items representing ownership may perhaps be made more vivid by taking a case of declining assets. Referring again to the last statement on page 86, let us suppose that a note for \$10,000 proves to be uncollectible, making it necessary to reduce the notes receivable and the total assets by \$10,000. Losses fall first upon the owners of the business, and we will therefore reduce the undivided profits by \$10,000 to make the two sides balance. A fire occurs and wipes out the plant entirely, causing a loss of \$500,000 in excess of the insurance. The total assets will be decreased by \$500,000, the \$102,560 still remaining in undivided profits will be wiped out, and surplus will be reduced by \$397,440, thus balancing the account.

Losses can continue to occur without affecting the outside creditors' position until the point is reached where capital stock, surplus, and undivided profits have been wiped out entirely. When that point is passed, the corporation is said to be insolvent. In precise terms a corporation is insolvent when its assets are less than its liabilities to others than the owners. When the assets are equal to or greater than the liabilities to outsiders, the corporation is solvent.

The value of stock. It has been customary to assign a definite value to each share of stock and to engrave this value on the stock

¹ The word "capital" is a much abused term, used in various senses which have become so entrenched in popular and even scientific language that we unfortunately cannot restrict it to one meaning without too violent a break with common usage. We have already defined capital as one of the factors of production (see Chapter II). We must now note that the term capital may be used also to designate what is in the text called the "net worth" of a corporation or of a partnership or individual. This should always be distinguished from the "capital stock" when it is not the same as the net worth. And finally we may use the same overworked term to designate all the resources at the disposal of a business enterprise, incorporated or not; that is, the total assets.

certificate. This value is known as the *par value*, and the par value, multiplied by the number of shares outstanding, gives the figure at which the capital stock is carried on the books of the corporation. While there are certain conveniences which attach to this custom, it has tended to confuse certain investors and others as to the true worth of a share of stock as reflected in either its actual or potential earning capacity. To engrave on a certificate of stock the words "ten shares, \$100 each" may be but the expression of a pious hope; the shares may be worth no more than the paper on which they are printed, or they may be worth much more than \$100 per share. The increasing use of stock with no par value assigned is evidence of the recognition of this fact.

A more reliable index of real worth than par value is found in the *book value*, which gives the accountant's view of the value of each share of stock. This is easily obtained by taking the value of the owners' equity and dividing it by the number of shares. According to the last balance sheet above the owners' equity was \$1,612,560, and there were 10,000 shares outstanding; each share must then have a book value of \$161.256. The book value fluctuates with the increase or decrease in the owners' equity; large earnings swell surplus and undivided profits and send the book value up, while losses, which of course must be borne by the owners, decrease the surplus and undivided profits and cause book value to decrease.

The book value will represent the real value of stock to the extent that the accountant's estimate of the value of the assets is correct. In some items there can be no possibility of error; if there is cash on hand or deposits in the banks, the accountant has no latitude. But in determining the figures at which the plant shall be carried in the statement and in estimating the proportion of bills and notes receivable which may eventually prove to be uncollectible there is room for considerable error. The conservative accountant has a tendency to undervalue the assets in order to be on the safe side; he may carry the real estate at the same figure for ten years or more, even though real estate values in general have increased greatly. On the other hand, the fly-by-night concern may greatly overvalue its assets in the attempt to bolster up the selling price of its stock, or to improve its credit standing, or possibly to conceal

transactions which have "milked" the corporation for the benefit of insiders. The book value, though in general much more significant than the par value, may be greater or less than the real value.

A third estimate of the value of a share of stock is its *market value*, the selling price of the share in the open market. This is based on the estimate of the dealers in the market of the earning power of the corporation. If there is an active market for the stock of a particular corporation, enough interest will be manifested by investors and professional traders in the affairs of the corporation so that the market value, aside from temporary speculative fluctuation in either direction, may be taken as a fairly trustworthy indication of its true value. It represents the appraisal of those who may be presumed most competent to judge.

Interest payments and dividends. We have already stated that the payment of interest on bonds is a part of the contractual relation between the corporation and the bondholder. If we assume the interest rate in our example above to be six per cent per annum, payable annually, the corporation evidently has a yearly charge of \$30,000 to meet. The payment of \$30,000 for this purpose would change our balance sheet as follows: (1) cash would be decreased by \$30,000, and (2) the item "interest due bondholders" would be wiped out entirely.¹ A simultaneous and equal change in both assets and liabilities leaves the totals equal, and no further change in the statement is required.

Let us assume next that the directors decide to pay a five per cent dividend to the stockholders. This means five per cent on the capital stock; it amounts to \$50,000, or \$5 per share of stock. Cash would be decreased by \$50,000, and undivided profits would be similarly decreased. The owners' equity is reduced to \$1,562,560, and if we calculate the book value of each share of stock we shall see that it is now \$156.256 (*i.e.*, \$1,562,560 divided by 10,000); in other words it has decreased by precisely the amount paid in dividends on each share of stock.

¹ From the fact that the item "interest due bondholders" in the statement of July 1, 1939, is \$30,000, the exact amount of the annual interest charge, we infer that this is the date on which the annual payment of interest is due. On any other date this item would be something less. Immediately after the payment, it is zero, as stated in the text; thereafter it increases gradually, as the liability "accrues," until it again reaches \$30,000 on the date when payment is due one year later.

Stock dividends. Suppose the Steel Manufacturing Corporation decides to declare a *stock dividend* of fifty per cent. Each shareholder will receive half as many shares as he now possesses. A, who owns 5,000 shares, will receive 2,500, and the other shareholders in proportion. It is quite evident that this does not disturb the assets in the slightest; so far as our statement is concerned all the changes will be among the liabilities. Capital stock will be increased to \$1,500,000 and the surplus of \$500,000 will be wiped out entirely. The net worth of the business is not altered, but the book value of each share of stock is decreased, for now we divide the net worth by 15,000 instead of by 10,000.

It is quite evident that the position of the stockholders has not changed at all so far as their claims upon the assets of the corporation are concerned. The 5,000 shares which A owned had a book value before the stock dividend of \$161.256 per share, or a total value of \$806,280; now he holds 7,500 shares, each having a book value of \$107.504, or a total value of \$806,280. The book value per share of stock has declined, but the total book value is not affected. In fact the stock dividend, except in a figurative sense, is not a dividend at all, but merely a transfer from surplus to capital stock.

Since the stock dividend makes no change in the position of the stockholder with respect to the corporation, it may be asked what purpose it can serve. Without attempting a complete answer, the two most common purposes may be cited. In the first place, it may happen that the market value of the stock is so far in advance of the par value that a share of stock is less easily traded in. Five shares which sell for \$100 each are much more convenient than one share which sells for \$500. The stock dividend may then be resorted to in order to increase the number of the shares and lower the value of each. But the chief motive is probably to avoid the appearance of unusually high profits and high dividend rates. A dividend of 20 per cent on a capital stock of \$1,000,000 is no greater than a dividend of 10 per cent on a capital stock of \$2,000,000, but it seems greater; the former may attract considerable notoriety, lead to the suspicion of monopoly, and invite competition, while the other passes by without exciting comment.

Watered stock. Stock is said to be "watered" when the corporation issuing it has not received full value for it. There are several different ways of watering the stock. One of the most common is to issue stock in payment for property or services which are overvalued. Suppose that \$100,000 worth of stock, par value \$100 per share, is sold to investors for cash, and that another block of \$100,000 is given in payment for a tract of land which, at the most liberal estimate, is worth not more than \$50,000. The statement would show assets of \$200,000 and liabilities of \$200,000, and the book value of each share would be \$100. A true record on the books of the corporation of the value of the assets would however record them as worth \$150,000 and would thus show a book value of only \$75 per share. A stock dividend without a sufficient surplus to justify it is a form of stock watering. In our example above (page 86) a stock dividend of a million dollars could have been declared only by raising the recorded value of the assets by \$387,440.

The essence of stock watering is overvaluation of the assets of the corporation, and it may be resorted to for any of the purposes stated in a previous paragraph dealing with book value. Like the legitimate stock dividend, its purpose may be to conceal large profits, though in this case it really conceals them instead of merely avoiding an appearance of high profits. Whatever its purpose, stock watering involves falsification of accounts and carries at least the presumption of dishonest intent.

The corporation's income account. The balance sheet or property account shows the condition of business at a given instant of time, but it does not give all the information that may be required regarding the condition of the business. Even a comparison of two such accounts for successive years, while more informing than a single balance sheet, does not tell the whole story. It does indeed show the changes in the various items in the balance sheet and it may record the profitableness of the business, but it does not tell us how the changes have come about or what has occasioned the profits or the losses. To complete the picture we must have another account, called the *Income Account* or the *Profit and Loss Statement*, which gives in condensed form the history of the operations of the business over the accounting period in question.

In order to illustrate the content, form, and meaning of a typical Profit and Loss Statement, let us assume that the City Grocery Company is dealing in eggs solely. On January 1 it had on hand eggs which cost it \$40,000. During the year it purchased eggs costing \$171,200, and it paid out in freight and cartage \$3,100 to bring these eggs to its warehouses. Its sales of eggs amounted to \$215,000, and at the close of the year it had on hand eggs which had cost \$42,300. Tabulating these figures we can get our first approximation of the operating profit of the business for the year, or the *gross profit*.

Net Sales		\$215,000
Deduct : Cost of Goods Sold		
Inventory Jan. 1, 1938	\$40,000	
Purchases	171,200	
Freight & Cartage Inward	<u>3,100</u>	
	\$214,300	
Less : Inventory Dec. 31, 1938	<u>42,300</u>	
Cost of Goods Sold		<u>172,000</u>
Gross Profit		\$43,000

The gross profit obtained by deducting from net sales the cost of the goods sold is not the final net profit actually earned ; other expenditures have been made and must be deducted. Expenditures for advertising, salesmen's salaries, commissions, and the maintenance of a delivery service are necessary to keep the merchandise moving from the warehouse to the consumer. Expenditures of this type are classified as *Selling Expenses* and must be deducted from gross profit. They cannot be included in the cost of goods sold, because, for the most part, they do not relate to any particular sale. Commissions could conceivably be deducted from selling prices, but convenience and uniformity demand that they be grouped with the other selling expenses.

In seeking the final net profit from operation it is also necessary to take into account another major group of expenses known as *General and Administrative Expense*. Such are the salaries to executives and to the office employees who keep the books and accounts, send out monthly statements, make collections, adjust complaints, supervise the purchasing of merchandise, etc. Office supplies of all kinds are consumed. Moreover provision must be

made for warehouses, offices, show rooms, etc., with all the necessary equipment, and depreciation on all these assets must be treated as an expense. Characteristic of these expenses is the fact that they relate almost entirely to periods of time rather than to the volume of sales. Depreciation occurs, and rent and taxes accrue at a uniform rate whether sales be large or small. Deducting selling expense and general administrative expense from gross profit gives *net profit from operation or net profit on sales*.

PROFIT AND LOSS STATEMENT

City Grocery Company

(For the year ending December 31, 1938)

<i>Net Sales</i>			\$215,000
<i>Deduct: Cost of Goods Sold:</i>			
Inventory Jan. 1, 1938	\$40,000		
Purchases	171,200		
Freight and Cartage Inward	3,100		
	<u>\$214,300</u>		
Less: Inventory Dec. 31, 1938	42,300		
Cost of Goods Sold		172,000	
<i>Gross Profit</i>			<u>\$43,000</u>
<i>Selling Expense:</i>			
Advertising	\$5,000		
Salesmen's Salaries	10,000		
Delivery Expense	<u>3,000</u>	\$18,000	
<i>General and Administrative Expense:</i>			
Salaries	\$11,200		
Office Expense	3,240		
General Expense	2,000		
Depreciation	1,500		
Taxes	<u>800</u>	\$18,740	
			<u>\$36,740</u>
<i>Net Operating Profit</i>			\$6,260
<i>Add: Other Income:</i>			
Interest Earned	\$300		
Rental Income	<u>240</u>		540
			<u>\$6,800</u>
<i>Deduct: Other Expense:</i>			
Interest Paid	\$1,100		
Loss on Sale of Real Estate	<u>700</u>		1,800
<i>Net Profit for the Year</i>			<u><u>\$5,000</u></u>
<i>Appropriation of Net Profit:</i>			
Dividends Declared	\$3,000		
Carried to Undivided Profits	<u>2,000</u>		
Total Net Profit			<u><u>\$5,000</u></u>

Logically the profit and loss statement should stop at this point, but in every business there are incidental receipts and payments, usually financial in character and not a part of the principal business of the corporation. Space may be leased to an outside tenant; interest may be received on investments; rent or interest may be paid; discounts may be earned or given, and the like. For want of a better title such income is called *Other Income* and such expenses *Other Expense*. Adjusting the net operating profit for these additional items gives the *Net Profit* for the year.

The Profit and Loss Statement of the City Grocery Company can now be presented in the usual form, as it appears on the opposite page. Note that in the first part (the trading section through gross profit) the figures depend upon the volume of merchandise handled, while in the remaining sections the figures are chiefly in proportion to the passing of time.

A comparative balance sheet for the beginning and ending of the accounting period is given below.

COMPARATIVE BALANCE SHEET

of the City Grocery Company

(As of the close of business December 31, 1937, and December 31, 1938)

	<i>Dec. 31, 1937</i>	<i>Dec. 31, 1938</i>
<i>Assets</i>		
Cash	\$5,000	\$4,200
Accounts Receivable	30,000	32,500
Inventory	40,000	42,300
Fixed Assets	25,000	31,000
	<u>\$100,000</u>	<u>\$110,000</u>
<i>Liabilities</i>		
Accounts Payable	\$26,500	\$34,500
Mortgage Payable	10,000	10,000
Capital Stock	50,000	50,000
Surplus	10,000	10,000
Undivided Profits	3,500	5,500
	<u>\$100,000</u>	<u>\$110,000</u>

From this we see that the net worth of the business has increased by \$2,000, but the balance sheet does not show us what brought this about. Nor does it tell us that \$3,000 was earned, declared, and paid out in the form of dividends. For this and other facts

we must have recourse to the income account. Neither the balance sheet nor the income account by itself is sufficient; together they show us in what respects the business has changed and what has caused these changes.

EXERCISES

1. Arrange the following items in the form of a statement of the assets and liabilities of a corporation : machinery, \$36,000; accounts receivable, \$12,500; capital stock, \$60,000; cash, \$7,500; bonds outstanding, \$50,000; notes payable, \$8,000; land and buildings, \$65,000; raw materials on hand, \$18,000; taxes payable, \$2,300; finished goods on hand, \$14,000; accounts payable, \$9,500; surplus, \$25,000; U. S. government bonds, \$5,000; undivided profits, \$6,200; office equipment and supplies, \$3,000.

2. There are 1,200 shares of capital stock outstanding. What is the par value of each share? the book value per share? NOTE: In this and each of the following five exercises, refer to the original statement as in Exercise 1.

3. A cash dividend of six per cent is declared and paid to the stockholders. Make the necessary changes in the statement. What is the book value per share now?

4. The company declares a stock dividend of twenty-five per cent. Make the necessary changes in the statement. What is the new book value per share? Stockholder A held twenty shares of capital stock before the stock dividend was declared. Calculate the book value of his holdings of capital stock before and after the stock dividend.

5. A fire breaks out and destroys the entire stock of raw materials and also a warehouse, valued at \$5,000, in which they were stored. These losses are not covered by insurance. Make the necessary changes in the statement. Is the company still solvent?

6. The management of the corporation decides that the valuation of certain assets should be reduced to reflect changed market conditions. The value of machinery is reduced by \$5,000; land and buildings by \$10,000; raw materials on hand by \$4,000; finished goods on hand by \$3,000. Make the necessary changes in the statement.

7. Make such changes in the balance sheet as would be necessary to illustrate stock watering.

8. The general ledger of the Apex Drug Company contains the following income and expense accounts with balances as indicated at December 31, 1938.

Purchases	\$28,700.00
Delivery expense	780.00
Depreciation of buildings and equipment	1,240.00
Salaries of office force	2,380.00
Inward freight and cartage	890.00
Sales (net)	39,480.00
Office supplies and postage	175.00

THE FORM OF THE BUSINESS UNIT 97

Advertising	340.00
Dividends declared	2,000.00
Interest paid	88.00
Taxes	427.00
Wages of sales force	4,160.00
Miscellaneous expenses (general)	325.00
Interest earned	35.00
Charitable contributions	25.00

The inventory of merchandise totaled \$12,500 on January 1, 1938, and \$14,260 on December 31, 1938.

From the above facts prepare a profit and loss statement for the Company for the year ended December 31, 1938.

VI

DIVISION OF LABOR AND THE CONTROL OF PRODUCTION

Coöperation. The key to an understanding of the modern industrial system is to be found in the idea of *coöperation*. By this is meant that the divers elements in society are consciously or unconsciously working together toward a common end.

There are historical examples of communities working together and consciously regulating their various relationships to each other. Such societies as the Oneida Community, the Amana Society, Brook Farm, and others were organized for this purpose. These examples of small isolated socialistic communities are however of only incidental interest to our present inquiry. Our task is to obtain a bird's-eye view of the present-day economic organization in order that we may see in what sense it is a coöperative system and may discover the motive forces which impel the various members of society thus to work together and the regulating forces which direct their efforts.

First of all we must acquire a clear notion of the meaning of coöperation and its several forms. Probably the earliest and certainly the simplest type of coöperation is found where men get together to perform a task which is too great for the strength of a single man. In former years the farmers of New England worked together at a barn-raising or in lifting a millstone into place. There are and always will be in everyday life countless examples of this form of coöperation, which we may call *simple coöperation*, and which is distinguished by the fact that the several coöperating workers are all performing the same kind of work.

Division of labor. Coöperation becomes far more important when it passes beyond this simple stage and leads to specialization, in which each worker confines himself to one kind of work, made possible because other kinds of work are cared for by other workers. Such coöperation is called *division of labor*. It is of varying com-

plexity according to the degree of specialization. The first steps occur when industry is separated into distinct trades. One man devotes his entire time to making shoes, another to raising wheat, another to making cloth, and so on through the various trades with which we are familiar. But the mere separation of trades, in which each worker performs all the several operations of his own trade, taking the raw materials and finally turning out the finished product, has only begun to exploit the possibilities of division of labor. Every trade is itself a complex of many distinct operations, and division of labor has not reached its final goal until each of these operations is made a separate task and assigned to a special group of workers. Although it did not originate there, the more complex division of labor is seen to best advantage in the modern factory, where it has been developed to its highest point. Here we see the extent to which it has been found possible to split a trade up into distinct minor operations, each performed by a different group of workers. In the making of a pair of shoes the soles are cut by one man, the various parts of the uppers by others, one part is stitched by still another, and so on. Complex division of labor may be carried on without power machinery, but its possibilities are usually greatly enhanced by the employment of machinery for many of the operations. In the making of the Goodyear welt shoes it is said that there are one hundred and eighty-five distinct operations, of which one hundred and fifty-seven are performed by machines.

Economic effects of division of labor. Simple coöperation is in fact so simple that it will require no further discussion. Our attention for the remainder of this chapter will be devoted primarily to division of labor as the really fundamental form of coöperation.

The predominant fact about division of labor is that it increases man's efficiency in production. In the first place, division of labor, by permitting each worker to specialize upon one trade or even upon a single operation in a trade, permits him to acquire a degree of skill and dexterity which would otherwise be out of the question. Constant repetition gives a manual dexterity which soon becomes automatic and almost unconscious. Just so far as it is necessary to think out the next move speed is hampered. Contrast the swiftness and deftness of the seamstress, whose needle moves in and out with

automatic precision, with the clumsy efforts of the amateur. Contrast the man who is at the same time a farmer, a carpenter, and a blacksmith, with the specialists in these trades. The carpenter instinctively reaches for the proper tool; the farmer has to make a decision, and when he has determined to use a saw, let us say, his hand lacks the cunning which constant use has given to that of the carpenter. The farmer who builds a shed or a barn or paints his house in his spare hours requires much more time than the professional carpenter or painter. The simplification of the individual operations in a factory makes it possible for the operator to acquire a degree of skill and dexterity in his given operation which often appears little short of miraculous.

Division of labor also makes possible the saving of time which by the unspecialized worker is wasted in putting away the tools required for one job and getting the tools for another. A certain amount of time is used in really getting started, in deciding where and how to begin on the new job. Even a skilled house painter spends some time in getting up momentum, but once the pace is set he can keep it up pretty steadily through the working day. When the day is divided among several different tasks, much time can be wasted in the motions of work without accomplishing a great deal. This is one of the handicaps against which farm labor struggles, because of the limited opportunity for division of labor on the farm.

When division of labor does not prevail, there is also the social waste which comes of leaving tools idle during a good part of the day. If analyzed, it can be seen that this is also a waste of time for the group as a whole. Assume that we have five individuals and five possible trades. If each individual follows all five trades, twenty-five sets of tools are necessary; if there is specialization, only five sets will be required, and the saving of time required for making tools releases labor for other employment.

Division of labor not only economizes the use of tools and machines, but it leads to the invention and utilization of complicated labor-saving machinery which would otherwise not be possible or economical. As the processes of any trade are subdivided, each operation becomes more and more simple, and sooner or later it becomes possible to turn over the simpler ones to machines,

thus setting free labor for other tasks and increasing the product of industry. The reduction in the physical labor of tending machines has made it possible to use women for a variety of tasks which men alone used to perform, and the decreased difficulty of the task itself has considerably shortened the time spent in acquiring the necessary skill to perform it. Contrast the former apprenticeship of seven years with the few weeks or days which seem to suffice now to learn how to operate a machine.

Finally with division of labor it becomes possible for a man to find the work for which he is best fitted and in which he can show the greatest dexterity and speed. It by no means follows that in our society this adaptation will always take place, and in many cases it is quite apparent that it has not taken place. There is a tendency to follow customary lines of work. The son of a banker may follow banking as the line of least resistance, even though his natural bent may be in the direction of medicine. There is many an automobile mechanic who would be a more useful member of society as a farm hand, and doubtless many a farmer who is wasting his talents on the farm. But there is at least the possibility of this adaptation today, whereas formerly, as on the manor, nearly every man had to be a farm laborer whether he would or not. As a special application of this principle, we have the inestimable advantage of giving full time employment to capable business leaders, whose important place in our economic organization we have already noted. If one will think of some of the great industrial leaders — Carnegie, Harriman, and Ford, to mention only three names — and try to imagine them in the manorial period in any rank of life he chooses, he can better appreciate the extent of the opportunity which the present organization of society gives to superior ability and the degree to which society profits from the exercise of such talent.

All of these results of division of labor promote production, by economizing labor, increasing its efficiency, and making more effective use of capital. With a given amount of human labor there emerges a richer flow of commodities and services for the satisfaction of human wants. Or we can look at it from another standpoint and say that division of labor makes it possible to turn out the

same quantity of goods with less expenditure of effort, thus leaving more leisure for the enjoyment of the products of industry. It should be noted however that not everything that can be said of division of labor is in its favor. It is objected, for example, that modern specialized labor is monotonous and deadening in its effects on the workers, that modern machine tending is so largely automatic that skill is at a discount and labor is kept at a dead level, that the highly specialized laborer is unfitted for work in another trade. There is doubtless some measure of truth in these charges, but they may easily be exaggerated. Labor on the medieval manor was hardly free from monotony, and the shorter working hours of modern industry probably turn the scale in favor of the present day as regards the deadening effect of labor. For machine labor it can be said that, automatic though it be, the demand for skilled labor is still active and that labor could scarcely be more nearly on a dead level than it was on the manor. The charge that specialized training today makes change of occupation difficult is the reverse of the truth. An operative in a watch factory could fit into a bicycle factory more easily than a house painter could become a blacksmith. On the whole such drawbacks as are inherent in the division of labor are certainly not sufficient to prevent its net effect being overwhelmingly on the credit side of the account.

Division of labor and the extent of the market. Whether the farmer builds his own barn in addition to farming or hires a carpenter to do it for him depends in part on conditions over which he, as an individual, has little or no control. Assuming that his true bent is farming, he will prefer to hire the carpenter, but if he cannot sell the wheat that he does not need for his own family, he will have to build his barn. If the carpenter cannot sell his services he is forced to raise his own wheat, even though he be an indifferent farmer. The possibility of exchanging the surplus which one individual produces is a necessary preliminary to division of labor, and as exchange becomes easier there is a tendency to subdivide the tasks and to proceed from simple to complex division of labor. On the manor division of labor was of the most rudimentary sort, because there was no market in which surplus products could be sold. When the towns began to develop and to provide a market

we find the beginning of specialization on the part of the town craftsmen and the country laborers and the exchange of the surplus of each. When the market consists only of the town and the countryside, the village cobbler has little use for assistance and can himself fabricate the whole shoe. If he does require one or two helpers, there will be little division of labor among them. Each will be able and will be required to make a whole shoe, although it may develop that one will be particularly skilful in a special part of the work and devote a large part of his time to that operation. But if the shoemaker is called upon to manufacture several thousand pairs of shoes a year, it will be economical for him to hire men to perform minor operations, for then he can provide them with continuous work. It would further be profitable for him to use machinery to perform some of the operations.

Territorial division of labor. The extent to which division of labor will be carried depends then on the extent of the market — on the volume of goods which consumers are willing to purchase. The demand for a particular product may increase with the gradual growth of the population or as the result of a change in popular taste. But in the past the most striking cause of increased demand for the product of a particular manufacturer or locality has been a widening of the market through improvements in the means of transportation; in other words, through the addition of a demand which was previously satisfied by other producers or not satisfied at all.

Certain regions are peculiarly fitted by natural endowment for the pursuance of certain occupations. The soil and climate of the American southern states make them particularly suitable for the growing of cotton; the coal and iron of Pennsylvania make specialization in iron and steel products natural; the fertile plains of the mid-western states offer unusual advantages for growing grain. The specialization in production which is thus dictated is called *territorial* or *regional division of labor*. Under almost any circumstances there will be some degree of specialization in the products of the region, but unless adequate transportation facilities exist between different regions each must be self-reliant to a great degree. It is a familiar fact that a century ago New England was in large

measure dependent on her own food supply, in spite of the fact that her soil has never been as productive as that of the West. Yet New England could not specialize in manufacturing until she could market a volume of goods far in excess of the needs of her own people and could satisfy her food requirements by grain from the West. The great farming regions of the West, on the other hand, would never have been able to specialize in agriculture except for the possibility of selling their products in New England and other regions and purchasing from New England and other regions their stock of manufactured goods. It was the development of the transportation system that made possible this interchange of goods and this territorial division of labor, to the great advantage of both regions.

Division of labor a form of coöperation. Production with division of labor is clearly coöperative. Laborers, employers, capitalists, and landowners are working together in the sense that each is in part dependent on the efforts of others. In the long chain of productive operations each producer is dependent on the one below him for his raw materials. The cotton manufacturer relies on the grower of cotton, the cotton middleman, the coal miner, the manufacturer of cotton machinery, and literally a host of others, to provide him with the wherewithal to conduct his operations. The agriculturist grows more wheat than he can use, confident of his ability to sell the surplus and with the proceeds secure the cotton cloth that he needs. Coöperation in production consists in the creation of a surplus and the exchange of that surplus for the products of others. The laborer working for wages just as truly exchanges his surplus as does the manufacturer who hires him, for the manufacturer pays him the present worth of his product in money instead of giving him a portion of the product to sell, as might be done, and as is sometimes done with farm labor.

This fact of coöperation is by no means of academic interest only; it is of great practical importance. In the past, material progress has followed in the steps of coöperation. Each one of the successive industrial stages which we have studied is marked by a widening of the coöperative area and by an increase in the output of goods. As in the past so in the future, one path of material prog-

ress lies in the direction of increased coöperation, of greater specialization both within the industry and between regions having natural aptitudes in the production of certain goods. Any artificial restraint therefore which hinders this specialization decreases by so much the effectiveness of labor.

We cannot leave this topic without emphasizing the reliance which individuals, communities, and nations place on the effectiveness of coöperation in supplying them with the goods in which they are deficient. England affords the classic example. Lacking in sufficient food supply and in raw materials such as cotton, copper, and many minerals, she has been able, through utilizing her deposits of coal and iron and her intelligent population, to produce a volume of manufactures great enough to supply her own needs and at the same time, by selling the surplus abroad, to procure food for her population and raw materials for her industries. Any large city of the present day is entirely dependent on outside sources of supply for food and raw materials and is therefore dependent for its very life on transportation facilities. The tying up of all the rail connections to New York would be a catastrophe of the greatest magnitude. It can easily be seen that the threat of a nation-wide strike on the part of the railway workers is a powerful weapon, and that it is powerful because of the degree to which we depend on coöperation in production.

The control of production : Conscious and unconscious control. Production as a whole presents the truly marvellous spectacle of millions of men engaged in thousands of different tasks, turning out goods varying from the most essential to the most frivolous, transporting these goods where they are most urgently wanted, financing their manufacture, shipment, and sale, and performing personal services of the most varied character. Each producer is apparently quite independent in his choice, yet each attempts to turn out exactly the amount of his particular product which can be sold, and with few exceptions the goods all seem to reach the consumers where and when they are wanted. This coördination of human activity cannot be the consequence of accident, nor is it merely a fortunate coincidence; there must be some control. Now such control may be of two types: (1) by legal authority restricting

the freedom of the individual producer, or (2) by economic forces operating under a régime of personal liberty. In a certain sense we may call the first conscious, the second unconscious, regulation. Of the two types of control it is evidently the second which generally governs in the modern world. But this has not always been so, nor is it true everywhere today. For example, it is a matter of common knowledge that in Russia there is legal regulation of all economic activities. Indeed at the present time (1939) there is a definite trend toward governmental control in other important countries, such as Germany and Italy.

Regulation by legal authority. For the sake of a historical background for our study of the modern regulatory system let us digress for a moment and revert to a time in England when it was widely held that regulation of individual activity should be a conscious and definite policy of the government. The motives for regulation were in part to protect the subjects against the dishonesty or oppressive power of other nationals or foreigners and in part to increase the prosperity and prestige of the nation.

Protection for the individual. We have already seen that in the period of the town economy the protection of the inhabitants of one town against those of another was, in the modern view, carried to extreme lengths. The fixing of the price of bread was a common practice, and in the same spirit laws were passed providing for the inspection of cloth and other articles. This meant the practical determination by authority of the type of cloth which could be made. Some of these regulations lasted well into the nineteenth century, when they were either repealed or ceased to be enforced.

Another form of regulation was interference with the freedom of movement of the individual and limitation of his choice of occupation. We have already referred to the way in which the manorial laborer was tied to the soil. In the course of time the absolute restraint on moving from place to place was relaxed, but it was still true in the seventeenth and eighteenth centuries that custom and law made the laborer, and particularly the poor laborer, relatively fixed to the place in which he was born.

In the reign of Elizabeth the law of apprenticeship was passed, which compelled any person who wished to follow a trade to serve

an apprenticeship of seven years. This simply followed the custom of the time and was intended to make the efforts of the guilds to ensure honest workmanship more fruitful, but it tended to solidify a custom which but for this law might more easily have been modified in the course of time. It is true that this law applied only to such trades as were in existence when the law was passed and that it applied only to towns, but municipal regulations made it difficult to follow any trade unless one had been an apprentice for the required length of time.

Wages also were regulated by authority. According to the act of Elizabeth referred to above and subsequent legislation, the justices of the peace in each jurisdiction were required annually to determine what wages should be paid. It may be said that their efforts were not entirely unwelcome to the laborers, for while the justices were undoubtedly inspired with conservative views, it is nevertheless true that they frequently raised wages. In a period when workers were scattered and lacking in political influence, concerted action on their part to raise wages was quite out of the question. Hence any consideration of their needs by an authoritative body was not without advantage to them.

Promotion of national welfare. The second motive leading to regulation was the desire to increase the prosperity and power of the state. The sixteenth and seventeenth centuries were periods in which the nationalistic spirit was manifesting itself to an ever increasing degree, and it was generally agreed that the first duty of a ruler was to expand his power, whether at the expense of foreign nations or of his own subjects. Inasmuch as it was further believed that national and private interests were usually incompatible, it is easy to understand the restrictions placed on the private individual.

In furtherance of the aim to promote the national prosperity and power much attention was given to the regulation of the nation's foreign trade. Armed conflict was an almost constant feature, and the employment of paid standing armies in place of the feudal levies of former times as well as the increase in the scope of the state's activities made the accumulation of a store of the precious metals highly desirable, for in time of war that state which was most plentifully provided with ready money enjoyed a

great advantage. This led naturally and almost inevitably to a belief that gold and silver were the most important forms of wealth, and measures were taken to conserve and increase the supply. The most obvious and the earliest of these measures was a prohibition of the export of the precious metals. This legislation was unsuccessful in operation and was soon seen to be unwise, as there were occasions on which the export of gold and silver might result in an augmented importation.

Public attention was then directed toward starting a flow of gold and silver toward England through the regulation of imports and exports of goods. It was held that if the total value of the exports was greater than that of the imports the balance would of necessity be paid in money. That condition was called a "favorable balance of trade," and an "unfavorable balance of trade" was said to exist when the value of the imports exceeded that of the exports and money flowed out.¹

In order to secure a favorable balance of trade and to avoid an unfavorable balance heavy duties were imposed on the import of manufactured goods or on goods which could be produced in England. In some cases bounties were granted on the export of English wares. The importation of raw materials which could be worked up in England was generally encouraged, for the value of the goods when exported would be greatly enhanced and would therefore help swell the value of the exports. So also the export of domestic raw materials which could be manufactured at home and bring a much greater price in the finished state was prohibited. This was true of wool in particular. In summary we may say that regulation was intended so to affect the course of foreign trade that England would import only raw materials, which could later be exported in a more valuable form, and would export only finished goods.

¹ It is not our purpose here to examine the merits of this theory, but there are so many misconceptions based on it even at the present day, and it is so widely held in one form or another, that it must be stated emphatically that neither an excess of exports nor an excess of imports can be considered as either favorable or unfavorable and that it is impossible over a period of time for a country to have either a continuous inflow or a continuous outflow of gold, unless one of the countries is a gold-producing country and is exporting the gold in the same fashion as the United States exports wheat. This subject will come up for full discussion in a later chapter

Mercantilism. The political and economic philosophy of the period, which embraces these regulatory provisions, is called mercantilism. The foregoing account makes no pretense of giving a complete picture of all the restrictions thus imposed upon the producers by the laws of England, which, by the way, were distinctly more liberal than those of most other countries. As compared with the conditions which prevailed in France for example English workers and traders enjoyed a great measure of freedom. England had free internal trade by the eighteenth century, whereas the internal trade of France was impeded by innumerable customs barriers, and in the period of most minute regulation the individual could not determine what he should produce, how much he might produce, where he might produce it, or to whom he might sell it.

Laissez faire: Reaction from government control. During the eighteenth century the conviction spread that there had been too much government, too much interference by government with private business, and there arose a strong reaction against the economic restrictions of mercantilism. *Laissez faire* — let things alone — became the watch cry of philosophers, economists, and statesmen. *Laissez faire*, in the first place, emphasized the importance of the individual and of individual welfare. Secondly it assumed that individual welfare and national prosperity are not incompatible. It went further and asserted that individual prosperity is at the very foundation of national prosperity; that the individual, if allowed to follow his own self-interest without external restriction, would tend to exert himself to the utmost in the line of endeavor for which he was best fitted and would therefore be a more productive and hence a more valuable member of society than if his occupation and his conduct in that occupation were prescribed for him by either law or custom. As each man followed his own interest, the clash of interests would lead to modifications and compromises which would result in approximate justice to all.

Translated into terms of concrete policy, *laissez faire* supported (1) free trade, as opposed to a protective tariff, (2) no attempt to fix or regulate prices or to regulate either the quantity or the quality of goods produced, (3) entire freedom on the part of the individual to enter any occupation he might choose, in any place he might think

advisable, (4) the determination of wages by unfettered bargaining between employer and employee, (5) the limitation of the functions of government to (a) defense and the maintenance of peace and order, (b) the dispensing of justice, (c) the establishment and maintenance of certain public works and institutions, such as the paving of streets, the improvement of harbors and means of communication, the care of education, etc., which cannot be made commercially profitable as private enterprises.

Economic freedom. The reaction against mercantilism spread rapidly in the eighteenth and nineteenth centuries, and the general acceptance, particularly in England, of the doctrines of *laissez faire* ushered in a period in which the dominant note was the endeavor so far as possible to give free play to the forces of individual initiative. Up to the present day, despite the fact that the scope of government regulation has at many points been broadened beyond the limits of extreme *laissez faire*, the conditions under which production operates have on the whole been those of economic freedom in contrast to legal authority. Taking a broad view of the present system, we note that it is founded on three major principles: (1) personal liberty, (2) private property, and (3) individual initiative and control of enterprise.

So generally is personal liberty taken for granted in many parts of the world today that we are in some danger of failing to appreciate its full significance. In truth personal liberty is a comparatively new thing, only recently gained as the climax of a struggle of centuries. Throughout most of the world's history freedom was a rare privilege enjoyed by the favored few. The masses of the people were slaves or serfs, more or less completely subject to the will of their owners or overlords. Even after slavery was abolished and the status of serfdom gradually liberalized and finally developed into comparative personal freedom, the legal restrictions upon economic activity which we have been studying continued. Real economic freedom is scarcely a century old.

In Chapter I, where property was defined and analyzed, the institution of private property was taken for granted. But this also is a remarkable and fairly modern institution. There was probably very little private property in ancient days, common

ownership by the tribe or nation or ownership by the chief or king being the rule. The feudal system was based upon ownership of at least the most important element of the wealth, the land, by the king or other ruler. Even today a certain part of the wealth of every community, more in some countries than in others, is owned in common as the property of government. But in most modern nations the great bulk of all the wealth is owned by individuals. Private property involves not only the right of the owner to enjoy all the benefits of his wealth but also the right to say who shall own it after his death; that is, the right of bequest or inheritance. This attribute of private property greatly increases its significance as one of the conditions which control production in the modern world.

The development of individual initiative and control of enterprise, succeeding the régime of control by legal authority, has already been sufficiently discussed. In spite of recent communistic and fascist reactions, it is today generally not the congress or the legislature, the town council or the gild authorities that say what sorts of goods shall be produced and what quantities of each, where they shall be marketed, and at what prices sold. Such matters are generally left to the personal judgment and free initiative of the men who direct modern industry, the farmers and miners, manufacturers and merchants, bankers and brokers; in short, the entrepreneurs.

The motive to productive effort. The goal of all industry is production, in order that human wants may be satisfied. Each producer is, under the institution of personal and economic freedom, at liberty to produce whatever he chooses, and the institution of private property permits him to have and dispose of what he produces. Thus is the desire for the satisfaction of wants liberated to become the motive force behind production, as it could never be without freedom and private property. The slave may cherish wants without result in increased application and diligence. He does as little as he can without incurring the wrath of the overseer, because he knows that there is small connection between his efforts and the satisfaction of his wants. A freeman on the contrary has the right to enjoy the fruits of his toil, and he is spurred on to efforts

more or less energetic and effective according to the urgency of his wants and his productive capacity.

But as we have seen, modern society is organized upon a co-operative basis, which means, in the first place, that one does not produce the things that will satisfy his own wants but must first exchange his own product for goods produced by others, and secondly that production must submit to an extremely complex organization controlled by certain individuals, the entrepreneurs. For the great mass of the workers, the connection between effort and satisfaction of wants, though it still exists, becomes less direct and less obvious. Nevertheless the basic motive remains; the wage earner is aware that the amount of his wages depends, in some measure at least, upon the efficiency of his productive efforts, and the entrepreneur thinks of the profits of his business. The modern motive to productive effort is thus frankly self-regarding, and modern industry proceeds along the lines of competition, which means in general that each is acting independently in his own interest without primary regard to the interests of others.

The rule of price. Such being the motive to production, and its direction being in the hands of the entrepreneurs, what is it that guides the judgment of the latter and tells them what goods to produce, what quantities of each, where to send them, and to whom to sell them?

The answer is price. Prices determine whether there will be profits or losses and so invite or repel the entrepreneur as he casts his eye over all the possibilities in the field of production. Wages, the price of labor, turn laborers from one industry to another. Price directs the flow of savings into new capital equipment for this industry or that. Price determines the amount of iron ore that will be mined, the amount of wheat that will be raised, whether wheat will be sent to one country or another. Price is so omnipresent and so subtle a force that few are fully aware of its significance or realize how completely it controls our economic life.

A simple illustration will make this clear. Ask any farmer what he would do if he knew or firmly believed that the price of wheat were going to double. Having every reason to anticipate greater profits, he would undoubtedly try to grow more wheat per acre and

plant some acreage to wheat that had previously been devoted to another crop or lain fallow. In his efforts to produce more wheat he might bid for laborers, offering them higher wages to induce them to leave their present jobs, and he might purchase new machinery, so stimulating the manufacture of farm machinery and directing capital into that line of industry. Conversely a fall in the price of wheat, threatening lower profits or even losses, will reduce its production, may lower farm wages, and tends to check the flow of capital into farm machinery. The farmers of this country rejoiced in the high prices of the World War. Profits were large and production expanded enormously to make up for the interruption to the flow of wheat from Russia, Australia, and other combatant countries. The return of peace brought with it an approach to normal conditions. Australian wheat could get to the market, and France could devote capital and labor to raising wheat. The price of wheat dropped, and American farmers found themselves in a desperate plight. The explanation was simple; the world production of wheat was too great to be sold at a profit by all of the farmers. The automatic check was working, painfully no doubt, in the direction of decreased production, the only way to raise the price to a remunerative level.

Price is thus the medium through which the consumer makes his voice heard, telling those whose judgment directs industry how much of each good to produce and whether any particular good shall be produced at all. If someone is willing to pay a sufficient price, anything conceivable will be produced, from bread and shoes to diamond shoe buckles. If for something else no one is willing to pay a price from which the producer can make a profit, then that thing will not be made. There is no production of the old-fashioned spinning wheels or of clothing of a bygone style or of gold horse shoes, because no one will pay the price. High prices yielding large profits in any particular line attract the attention of the entrepreneurs, and they direct land, capital, and labor into that branch of production. Falling prices and smaller profits are the automatic checks to overproduction.

Competitive price is actually determined by bargaining between sellers and buyers, producers and consumers. Each consumer of

wheat is striving to purchase his wheat at the lowest possible price, while each seller strives to sell for the highest possible price. The only restraint imposed is that of self-interest. A seller hesitates to be too grasping lest he lose his trade to another seller; a buyer fears to hold off for a low price too long lest he lose the opportunity to buy at any price. If the existing stocks appear deficient, this bargaining of sellers and buyers tends to a rise in price and economy in present use with more held back for future needs. The opposite results follow the fall in price which is induced by abundant present stocks. And finally the high or low price encourages or discourages future production to replace the present stocks.

Taking stock of our economic organization. We have now arrived at the conclusion of the first part of our study. At the outset the reader was reminded that man differs from the lower animals fundamentally in the infinite number and variety of his wants and in the extraordinary range and intensity of his activities in the quest of their satisfaction, and Part I of this book has sought to present a sketch of the economic organization which has been built up for this purpose. The central theme, giving coherence and meaning to the infinitely complex picture, is that modern economic life is coöperative, based upon division of labor, individual freedom, private property, and the control of production and distribution by price. We have called this "unconscious" coöperation, in that the agents of production are following their own separate interests and are rarely aware of the fact that they are engaged in coöperative enterprise. Through examination of earlier economic systems of a different sort and by the aid of a brief glance at the evolutionary process by which the present system developed, it has become evident that the latter is more effective in the production of wealth and services for the satisfaction of human wants than any system which has preceded it in the history of mankind.

It must not be presumed however that the productive system works perfectly. No human institution is perfect. The critical student of current affairs has frequent occasion to observe that the productive machine does not run with perfect smoothness. He may note an occasional breakdown in the transportation system, a disturbance in the mining industry which interrupts the normal

flow of coal from the mines to the consumers, a strike or lock-out in a factory, or frenzied speculation in land forcing real estate prices up to levels which cannot be maintained. It is evident that to many a person the alleged freedom to choose an occupation, to produce what he wishes, and to have and enjoy the product of his industry seems more imaginary than real. It is certain that not all the men whose judgment guides the industrial machine are the ones best fitted for their posts. Capital does not instantly and without exception distribute itself among the several industries in perfect adjustment to their respective needs. There is obvious maladjustment of the labor force, and no one can claim that every parcel of land is always put to its most productive use. Competition appears not always to apportion rewards according to productive contribution but sometimes to favor the merely shrewd or the actually dishonest. Private property, in company with free competitive industry and reinforced by inheritance, occasions, or at any rate permits, great inequality in the distribution of the products of industry and of the capital instruments from which further products are to flow.

At distressingly frequent intervals have come business crises, which are marked, on the one hand, by widespread unemployment and acute want and, on the other hand, by the inability of producers to sell their products except at heavy loss. Drastic curtailment of production, rapidly falling price level, bankruptcies, and business distress are other characteristics of such periods.

It must also be evident to anyone who looks beneath the surface that, when government authority was displaced by free enterprise under the control of price as the directing force back of production, the control of production was actually given over to the consumers, or more specifically to those who have the wherewithal to pay for the things they want. As has been observed, nothing will be produced, no matter how beneficial it might presumably be to mankind, unless people are able and willing to pay a price equal to its cost of production, and any conceivable thing will be produced, no matter how frivolous or even positively harmful, so long as someone is prepared to pay for its cost. Production is thus directed, not toward the maximum of things most beneficial to mankind

on the basis of some ethical or moral standard, but toward the things which are wanted by those who can pay for them. It is this which gives special significance to the inequality in the distribution of wealth which appears to be an inevitable concomitant of the present economic system. Those who have abundance do not merely receive a greater share of the product of industry; they have a disproportionate voice in saying what things shall be produced. The machinery of production is directed to the turning out of steam yachts and race horses, fine silks and costly wines, even though there be those who are in need of the food and clothing which might otherwise be produced in greater abundance; land is set aside for private estates and hunting preserves which might otherwise be devoted to the production of a more abundant flow of food and other necessities. Moreover the decisions of a comparatively small group as to whether to spend or invest their incomes may have important reactions upon the whole economic structure.

Why our economic organization thus falls short of perfect functioning is a question whose answer cannot be attempted at this point in our study. There are however two causes so basic and so generally evident as to warrant mention. In the first place, unregulated competition may at certain points degenerate into monopoly, and it is obvious that when monopoly takes the place of competition the public no longer has assurance that its interests are being safeguarded by the operation of an unregulated price system.

In the second place, there is in the economic world an evident lag between cause and effect. To illustrate, a decline in the price of wheat, caused by a decrease in the demand for wheat, will eventually (other things being equal) cause a decline in the area of land devoted to raising wheat and in the quantity of wheat placed on the market. It may however have surprisingly little immediate effect because of the commitments which farmers have already made, and over a short period of years its effect may even be an increase in the volume of wheat to be marketed. An individual farmer finds that to pay taxes, interest, insurance, and other costs he must sell more bushels at the lower price. The farmer who has just been making a living at the old price should, to avoid farming at a loss, turn now to a more profitable occupation, but transfer of

occupation is not effected easily or quickly. He may think it better for a time to raise enough food to maintain life, to try to stave off his creditors, and to wait for better prices.

To the natural slowness of adjustment we find added such artificial impediments to price adjustment as the maintenance of a fixed price policy by a monopoly, or the stubborn resistance to wage cuts which is characteristic of trade unions. These causes provide opportunity for even more serious maladjustments to make themselves manifest in the economic structure.

Finally the reader should perhaps be reminded that the foregoing condensed summary has of necessity been limited to only the broad lines of the picture, omitting numerous exceptions and qualifications. One needs to be on his guard against the impression that there prevails in the world today a complete system of free enterprise, with automatic price control of production and distribution, and that this system is firmly established on a presumably permanent basis. In important parts of the world — Russia, Italy, and Germany, for example — this is certainly not the case. Even in Great Britain, America, and other democratic regions, there are numerous regulatory provisions, which limit more or less severely the freedom of private initiative. In later chapters of this book these regulations and the problems involved therein will be examined. For the present let us recognize that in America and many other parts of the world the prevailing economic system is basically one of personal liberty and private property, but with very many restrictions and regulations, and that there is present in other parts of the world a strong current toward extremes of public control.

Shall the system be scrapped? There are those to whom these weaknesses of the present industrial régime loom so large that their patience is exhausted and they cry: away with the whole system! Industry is going to rack and ruin in the hands of irresponsible and incompetent entrepreneurs; let us substitute government ownership. Competition is nothing but legalized selfishness, in which the race is to the unscrupulous and grasping, not to the honest and efficient; let us therefore substitute the real coöperation which socialism offers. Worst of all is the intolerable inequality in

the distribution of the good things of life; let us therefore prohibit inheritance, let us have government ownership of the capital instruments of production, or better still let us abolish private property entirely and entrust all wealth, along with the direction of industry, to the socialistic state.

What the present system accomplishes. Now the time to pass judgment upon proposals to alter radically the existing economic system is not at the beginning of our study of economics but at its close, when we shall better understand the details of the present system and the possibilities of proposed substitutes. What our present purpose requires is to attain a clear notion of the general outline of the existing system and its operation. The essential foundation of the philosophy of economic freedom, with its corollaries of private property, individual control of industry, and competition, is that, by leaving the several individuals free to act in their own interests, the maximum total production will be obtained and to that extent at least the interests of the whole community will be served. As was pointed out in the first chapter, the welfare of mankind depends first upon the total amount of production and second upon its distribution. Any "reform" which aims to improve distribution must see to it that the effectiveness of production is not incidentally impaired, lest the remedy be worse than the disease.

There can be no doubt that the system of economic freedom does offer a powerful motive to production and does spur men on to energetic effort. Equally without doubt is it that economic freedom is an incentive to leadership. Executive ability of a high order is by no means over-abundant, and it is indispensable that such leadership be found out and placed at the disposal of industry. He is most successful who not only perfects the internal organization of his own business but who makes the most perfect adjustment of his business with other types of business round about him. The present system does on the whole call forth efficient leaders and make them coöperate. Again there can be no doubt that economic freedom offers a powerful stimulus to saving. The motives for saving are varied, but they are particularly strong in a system of private property. Whatever these motives may be,

we find that most of them are conditioned on the responsibility of the individual for his own present and future welfare, which is a corollary of freedom. The present system thus promotes the accumulation of capital needed for production.

Again it is certain that the automatic control of industry through competitive prices does on the whole compel producers to bring forth the goods the people want and to deliver them to the consumers at about what it costs to produce them. This is an accomplishment not to be lightly cast aside, even though the direction of industry may not perfectly conform to some preconceived standard of ethics or morals. And finally the present system of distribution, in spite of all exceptions, does bear a very definite relation to the productive contributions of the several individuals.

The present conclusion. It will be noted that the statements of the foregoing section leave ample room for all the exceptions to which attention has previously been called and for the perfectly evident conclusion that the present system falls short of perfection. It is therefore to be presumed that the present system is capable of improvement. In fact the system is already the result of first a movement away from government control and later a material reaction from the extreme of *laissez faire*. That further modifications of the present economic system are to come is a reasonable, indeed an inevitable, inference. That the whole system will be scrapped and something quite different substituted is most unlikely, just as unlikely as that it will be crystallized and perpetuated in exactly its present form. We shall give thought to these questions in the concluding chapters of this book. We have now reached the point in our study at which we can at least appreciate the well-nigh infinite complexity of our economic organization and can recognize that he who would attain to such understanding of its nature as to be able to pass wise judgment upon it has set himself no mean task. At the present stage of our study it would therefore be most unscientific either to jump to the conclusion that the imperfections of the present system warrant its discard in favor of something radically different or to conclude from the undoubted services to mankind of the present system that it is perfect and should be defended against any change. In the chapters

now before us we shall seek a more complete and intimate acquaintance with the economic system and shall thereby at least approach nearer to that knowledge and understanding which alone can qualify one to sit in judgment upon it.

SUGGESTIONS FOR FURTHER READING ON PART I, THE ECONOMIC ORGANIZATION

I. GENERAL ECONOMICS TEXTS

- SLICHTER, S. H. *Modern Economic Society*. New York, 1931
- ELY, R. T., and others. *An Outline of Economics*. Sixth Edition. New York, 1937
- EDIE, L. D. *Economics: Principles and Problems*. Second edition. New York, 1932
- GARVER, F. B., and HANSON, A. H. *Principles of Economics*. Boston, 1937
- SPAHR, W. E., and others. *Economic Principles and Problems*. Third edition. New York, 1936
- MEYERS, A. L. *Elements of Modern Economics*. New York, 1937
- ROLL, E. *Elements of Economic Theory*. London, 1937
- DEIBLER, F. S. *Principles of Economics*. New York, 1929
- MARSHALL, A. *Principles of Economics*. Eighth edition. London, 1920
- ORTON, W. A. *Prelude to Economics*. Boston, 1932
- FISHER, I. *Elementary Principles of Economics*. New York, 1912
- Economics and Social Institutions*. Boston, 1937-1938
- Vol. I. Modlin, G. M., and DeVyver, S. T. *Development of Economic Society*, 1937
- Vol. II. McIsaac, A. M., and Smith, J. G. *Introduction to Economic Analysis*, 1937
- Vol. III. Modlin, G. M., and McIsaac, A. M. *Social Control of Industry*, 1938
- Vol. IV. Luthringer, G. F., Chandler, L. V., and Cline, D. C. *Money, Credit, and Finance*, 1938
- Vol. V. Dell, B. N., and Luthringer, G. F. *Population, Resources, and Trade*, 1938
- Vol. VI. McCabe, D. A., and Lester, R. A. *Labor and Social Organization*, 1938

II. FUNDAMENTAL CONCEPTS

- FRASER, L. M. *Economic Thought and Language*. New York, 1937
- CANNAN, E. *Review of Economic Theory*. London, 1930
- ROBBINS, L. *Nature and Significance of Economic Science*. New York, 1935
- FOWLER, R. F. *The Depreciation of Capital*. London, 1934
- FISHER, I. *The Nature of Capital and Income*. New York, 1906
- *The Theory of Interest*. New York, 1930
- COMMONS, J. R. *Institutional Economics*. New York, 1934

- McCONNELL, D. W., and others. *Economic Behavior*. Revised edition. Boston, 1939
- EDIE, L. D. *Principles of the New Economics*. New York, 1922
- DICKINSON, Z. C. *Economic Motives*. Cambridge, 1922

III. EUROPEAN ECONOMIC HISTORY

- ASHLEY, W. J. *The Economic Organisation of England*. London, 1914
- *An Introduction to English Economic History and Theory*. New York, 1910
- EHRENBERG, R. *Capital and Finance in the Age of the Renaissance*. New York, 1928
- HECKSCHER, E. F. *Mercantilism*. 2 volumes. New York, 1935
- NUSSBAUM, F. L. *A History of Economic Institutions of Modern Europe*. New York, 1933
- FAY, C. R. *Life and Labor in the Nineteenth Century*. New York, 1933
- KNOWLES, L. C. A. *The Industrial and Commercial Revolutions in Great Britain during the Nineteenth Century*. London, 1921
- *Comparative View of Dominion Problems*. London, 1930
- *Economic Development in the Nineteenth Century: France, Germany, Russia, and the United States*. London, 1932
- DAY, C. *History of Commerce*. New York, 1922
- *Economic Development in Modern Europe*. New York, 1933
- HAMMOND, J. L. and B. *The Rise of Modern Industry*. Fifth edition. New York, 1937
- SINCLAIR, H. M. *A Preface to Economic History*. New York, 1934
- OGG, F. A., and SHARP, W. R. *The Economic Development of Modern Europe*. New York, 1926
- LIPSON, E. *The Economic History of England*. Volume I, *The Middle Ages*. New York, 1929
- Volumes II and III, *The Age of Mercantilism*. New York, 1931
- HEATON, H. *Economic History of Europe*. New York, 1936
- *The Yorkshire Woollen and Worsted Industries*. Oxford, 1920
- SALZMAN, L. F. *English Industries of the Middle Ages*. London and New York, 1923
- KRAMER, S. *The English Craft Gilds*. New York, 1927
- CLAPHAM, J. H. *The Economic Development of France and Germany, 1815-1914*. London, 1921
- *An Economic History of Modern Britain*, 3 vols. New York, 1927, 1932, 1937
- SEE, H. *Economic and Social Conditions in France during the Eighteenth Century*. New York, 1931
- THOMPSON, J. W. *An Economic and Social History of the Middle Ages*. New York, 1928
- PIRENNE, H. *Economic and Social History of Medieval Europe*. New York, 1937
- HUTCHINS, B. L., and HARRISON, A. *A History of Factory Legislation*. Westminster, 1911
- WEBB, S. and B. *The History of Trade Unionism*. New York, 1920
- JOHNSON, E. A. J. *Some Origins of the Modern World*. New York, 1936

IV. AMERICAN ECONOMIC HISTORY

- SHANNON, F. A. *Economic History of the People of the United States*. New York, 1934
- FAULKNER, H. W. *American Economic History*. New York, 1931
- BOGART, E. D. *Economic History of the United States*. Revised edition. New York, 1938
- CALLENDER, G. S. *Selections from the Economic History of the United States*. Boston, 1909
- COMAN, K. *Industrial History of the United States*. New York, 1905
- CLARKE, V. S. *A History of Manufactures, 1607-1860*. Washington, 1916
- DURAND, E. D. *American Industry and Commerce*. Boston, 1930
- HICKS, F. C. *High Finance in the Sixties*. New Haven, 1929
- KIRKLAND, E. C. *History of American Economic Life*. New York, 1932
- MYERS, G. *History of the Great American Fortunes*. New York, 1936
- COLE, A. H. *Fluctuations in American Business, 1790-1850*. New York, 1935
- GRAYSON, T. J. *Leaders and Periods of American Finance*. London, 1932

V. CORPORATIONS AND INVESTMENT

(See also suggested readings at end of Part VII)

- FIELD, K. *Corporation Finance*. New York, 1938
- HOOGLAND, H. E. *Corporation Finance*. New York, 1938
- STEVENS, W. M. *Financial Organization and Administration*. New York, 1934
- HANEY, L. H. *Business Organization and Combination*. New York, 1934
- BISHOP, A. L. *Financing of Business Enterprises*. New York, 1928
- DEWING, A. S. *Financial Policy of Corporations*. Revised edition. New York, 1934
- *Corporation Finance*. New York, 1931
- *Corporate Promotions and Reorganizations*. New York, 1914
- BURTCHETT, F. B. *Corporation Finance*. New York, 1934
- Twentieth Century Fund. *The Securities Market*. New York, 1935
- *Big Business, Its Growth and Its Place*. New York, 1937
- MEAD, E. S. *Corporation Finance*. New York, 1930
- LYON, H. *Corporations and Their Financing*. Boston, 1938
- WHITE, P. *Business Management*. New York, 1926
- LINCOLN, E. E. *Applied Business Finance*. Chicago, 1924
- GERSTENBERG, C. W. *Financial Organization and Management of Business*. New York, 1924
- CHAMBERLAIN, L., and HAY, W. W. *Investment and Speculation*. New York, 1931
- GRAHAM, B., and Dodd, D. L. *Security Analysis*. New York, 1934
- LYON, H. *Investment*. Boston, 1926
- JORDAN, D. F. *On Investments*. New York, 1934
- LAGERQUIST, W. E. *Investment Analysis*. New York, 1924
- KIRSHMAN, J. E. *Principles of Investment*. Chicago, 1933
- SAKOLSKI, A. M. *Elements of Bond Investment*. New York, 1921

- FOWLER, J. *American Investment Trusts*. New York, 1928
- Royal Institute of International Affairs. *The Problem of International Investment*. New York, 1937
- MADDEN, J. T., and NADLER, M. *Foreign Securities*. New York, 1929
- LEWIS, C. *America's Stake in International Investments*. Brookings Institution, Washington, 1938
- MADDEN, J. T., and others. *America's Experience as a Creditor Nation*. New York, 1937
- ROSE, D. C. *A Scientific Approach to Investment Management*. New York, 1929
- SMITH, E. L. *Common Stocks as Long Term Investments*. New York, 1925
- PRIME, J. H. *Analysis of Industrial Securities*. New York, 1935
- GRAHAM, B., and DODD, D. L. *Security Analysis*. New York, 1936

VI. STATISTICS AND GENERAL STATISTICAL SOURCES

- SMITH, J. G. *Elementary Statistics*. New York, 1934
- CROXTON, F. E., and COWDON, D. J. *Practical Business Statistics*. New York, 1934
- MILLS, F. C. *Statistical Methods Applied to Economics and Business*. New York, 1938
- LOVITT, W. V., and HOLTZCLAW, H. F. *Statistics*. New York, 1929
- BOWLEY, A. L. *Elements of Statistics*. London, 1920
- CHADDOCK, R. E. *Principles and Methods of Statistics*. Boston, 1925
- CRUM, W. L., PATTON, A. C., and TEBBUTT, A. R. *Introduction to Economic Statistics*. New York, 1938
- SECRIST, H. *An Introduction to Statistical Methods*. New York, 1917
- *Statistics in Business*. New York, 1920
- FISHER, I. *The Making of Index Numbers*. Boston, 1922
- MITCHELL, W. C. *The Making and Using of Index Numbers*. Bulletin of the United States Bureau of Labor Statistics, No. 284. October, 1921
- MUDGETT, B. D. *Statistical Tables and Graphs*. New York, 1930
- SCHMECKEBIER, L. F. *The Statistical Work of the National Government*. Baltimore, 1925
- Statistical Abstract of the United States*. Bureau of Foreign and Domestic Commerce
- Abstract of the Census of the United States*. Bureau of the Census

VII. ACCOUNTING

- HUSBAND, G. R., and THOMAS, A. E. *Principles of Accounting*. New York, 1935
- PORTER, C. H., and FISKE, W. P. *Accounting*. New York, 1935
- KESTER, R. B. *Accounting Theory and Practice*. New York, 1930
- HATFIELD, H. R. *Accounting*. New York, 1927
- COLE, W. M. *Fundamentals of Accounting*. Boston, 1921
- FINNEY, H. A. *Introduction to Principles of Accounting*. Revised edition. New York, 1936

VIII. ECONOMIC JOURNALS

- The American Economic Review.* Cambridge, Mass.
The Quarterly Journal of Economics. Cambridge, Mass.
The Journal of Political Economy. Chicago, Ill.
The Economic Journal. London
The Annals of the American Academy of Political and Social Science. Philadelphia, Pa.
The Political Science Quarterly. New York
The Journal of the American Statistical Association. Concord, N. H.
The Economist. London.
Econometrica. Menasha, Wis.
The Commercial and Financial Chronicle. New York
The Review of Economic Statistics. Cambridge, Mass.
The Federal Reserve Bulletin. Washington, D. C.

PART II

THE FORCES DETERMINING PRICE



VII

PRICE AND THE CONSUMER

Introduction. In harmony with the title and purpose of Part I of this book, the treatment thus far has been descriptive and historical rather than analytical and critical. We have not stopped always to go to the bottom in search of causes and explanations, nor have we gone far into technical details. We have not attempted to answer all of the many questions that have presented themselves from time to time or the many other equally important questions that have doubtless occurred to the reader. Our purpose has been to lay a foundation for the study of economics in a broad knowledge of the economic organization and its development. From now on our study will be more intensive, more analytical, more critical. We shall give more thought to the discovery of the principles which explain what goes on in the economic world, and we shall seek answers to a multitude of searching practical questions which are being asked about our economic organization and its operation.

The role of price: Price controls the material enjoyments of life. In spite of private property and personal freedom, the modern economic system is one of universal albeit unconscious co-operation, made possible by the control over production and distribution which is exercised by price. Realization of the significance of price is the point of departure for our further inquiry into the nature of the economic world in which we live.

With virtually universal division of labor, economic life centres about the exchange of goods and services, through the medium of money. With few and trifling exceptions, whatever anyone produces must be sold for money, and, by the same token, whatever anyone acquires must come by purchase with money. Prices therefore, together with income, control each person's enjoyment of the material things of life. The rich man has many enjoyments because his large income permits him to pay the prices demanded

by those who furnish commodities and services. The poor man has less to enjoy, because he cannot pay the prices. Particular prices, in their relation to each other, control the choices of the things which anyone may obtain to satisfy his wants. When railroad fares are high, many persons are compelled to cut down on their pleasure travel. On the other hand, when coal prices soar, we do not reduce much our consumption of coal; we are forced to make the saving somewhere else, on some article whose use we can better afford to give up or to reduce. A change in the price of a necessity may thus affect the consumption of comforts and luxuries more than the consumption of the necessity itself.

Among comforts and luxuries the choice is controlled by price. Polo is the sport of the few, baseball, of the many; not because baseball is inherently a more popular game than polo, but because of the difference in the prices of the respective equipments. When the genius of Henry Ford brought down the price of his product, the automobile ceased to be the rare luxury of the rich and became the common possession of those of moderate means. A generation ago attendance upon the theatre was a luxury, to be enjoyed almost exclusively by the well-to-do and only occasionally indulged in by the poor. Since the moving picture industry has made it possible to see an interesting performance for as little as fifteen or twenty cents, this enjoyment has become the common possession of the poor. Indeed through changes in price the classification of enjoyments into necessities, comforts, and luxuries (a loose classification at best) is subject to constant revision. There was a time when good woollen and cotton textiles were destined solely for the consumption of the rich and well-to-do. Today the prices of such fabrics are so low that they are used by everybody and regarded as among the prime necessities of life.

Even among necessities there is often choice between substitutes, the decision being determined by their relative prices. Fuel for cooking is a household necessity, but whether the fuel shall be wood, coal, oil, artificial gas, or natural gas is a problem to be determined according to the relative prices of these various fuels. There have been occasions when the farmers of the middle west burned corn for heating their houses and cooking their food, a choice forced upon

them by the relative prices of corn and other fuels. Ordinarily these prices forbid the use of corn as fuel.

From the viewpoint of consumption price thus appears as an all-powerful arbiter, deciding for us what clothing we shall wear, what food we shall eat, what sports we shall enjoy, whether we shall travel or stay at home, whether we shall have many pleasures or few, whether we shall live well or poorly.

Price from the producer's viewpoint. Most of us are producers as well as consumers. To the one who works for salary or wages, the price which his labor commands is the main factor determining his income and, jointly with the prices of goods in general, determining the total amount of his material enjoyments. Thus income itself is the result of price. To the manufacturer, the merchant, and the business man generally, everything centres about price. The prices of labor (wages), of materials, and of capital (interest) determine his costs; the prices of his product determine his income. Upon these prices depends the result of his enterprise — whether profit or loss. In all his plans, in all his operations, the business man is constantly thinking of price; everything finally reduces itself in his mind to terms of money.

A pecuniary system. This pecuniary character of modern life, while in a sense self-evident, is not always given its due place in our thinking. In order to arrive at fundamentals, we are frequently reminded that money is not the aim of economic exertion but is only a medium, serving the purpose of passing goods from producer to consumer with a facility not possible under a régime of barter; the ultimate social purpose of production is not the obtaining of money, but the production of goods to satisfy human wants; and the acquisition of money (except to the traditional miser) is not for its own sake, but simply as a means to the possession of the things that satisfy wants.

These are indeed profound truths. But they are not the terms in which the business man thinks. To the shoe manufacturer, his factory is operated, not to the end that the community may be comfortably shod, but that profits may accrue to himself. His payment of wages and other costs are not thought of as the means whereby his laborers and others may obtain satisfaction of their wants; they

are considered as payments of money, necessary to the conduct of the business — a necessary evil, if you like. The disposal of his product is thought of in terms of the money brought in, which he regards as the true purpose of his enterprise. Money costs and money income, money gains and money losses; these are the ultimate considerations in the mind of the business man. In a certain very true sense the shoe manufacturer's goal is not making shoes, but "making money." This is the business man's point of view. And the business man is the centre of the modern economic system; the business point of view gives character to all economic life. The modern system is in truth a pecuniary system, a money economy, centred about the exchange of goods and services for money prices.

The behavior of price: A bird's-eye view. If now, convinced of the importance of the subject of price, one turns his attention to actual prices and their behavior, the first impression is likely to be one of utmost confusion. An infinite number of articles is on sale, each at its own particular price, from the one cent postage stamp and the five cent lead pencil to the million dollar steam yacht. One kind of labor may be obtained for a dollar or two a day. Other labor commands \$50,000 a year. Ten dollars will buy a suit of clothes in one store; the exclusive tailor in the next block may get \$100. The shelves of the ten cent store are loaded with thousands of articles each priced at a dime.

To make confusion worse confounded, it appears that there is no stability in the realm of price. Prices are constantly changing. In the fall of 1929 the retail price of anthracite coal in New York was fifteen dollars a ton. By 1933 it fell below twelve dollars. Early in 1930 the average retail price of sirloin steak in fifty-one large American cities was about fifty cents a pound. In these same cities in 1933 sirloin steak sold on the average for less than thirty cents a pound. In New Haven the service station price of regular gasoline per gallon (including tax) was \$0.18 in May, 1929, \$0.125 in May, 1933, \$0.157 in May, 1935, \$0.175 in May, 1938.

Such price changes are fresh in the memory of the present generation; but price changes have been going on through all the centuries of modern history. For example the price of wheat in England in the fourteenth century was on the average about twenty-five cents

a bushel in United States money; in the year 1527 the price suddenly more than doubled. In the fall of 1929 wheat was in the neighborhood of \$1.35. Here are a few random examples from American price history: Pine boards cost twice as much in 1891 as in 1860; white oak boards cost more than three times as much. The price of sugar on the other hand declined by 1891 to a little more than half its price in 1860. Similarly salt in 1891 cost just about half what it cost in 1860. The price of milk was the same in 1891 as in 1860. During the next generation prices of most things increased. For example the wholesale price of salt rose from seventy-nine cents a barrel in 1891 to \$2.70 in 1921, three and a half times what it was in 1891. Eggs doubled in price from 1891 to 1921. White pine boards cost nearly five times as much in 1921 as in 1891. Granulated sugar on the other hand increased moderately, the wholesale prices being 4.7 cents a pound in 1891 and 6.2 cents in 1921. Some few things declined in price. Coffee for example cost 16.7 cents per pound wholesale in 1891 and 7.2 cents in 1921, a decline of more than half.

Here are the quotations of cotton on the New York market on June 21 of certain selected years from 1905 to 1935, showing a remarkable succession of changes up and down:

PRICE OF COTTON (MIDDLING UPLAND) IN NEW YORK

1905	\$0.09	1917	\$0.257	1924	\$0.2965
1910	0.153	1920	0.3875	1932	0.0535
1915	0.096	1921	0.1105	1935	0.1185
				1938	0.0871

Wholesale prices are generally much more variable than retail prices. The most startling price changes occur on the speculative markets, such as the wheat market and above all the stock market. Here there are "booms" when most prices soar, and "panics" when most prices come tumbling down as in the great debacle of 1929. One fateful day, October 24, 1929, the prices of industrial stocks on the New York Stock Exchange declined on the average about 13 per cent.

Individual prices sometimes appear to fluctuate without reference to each other. There is nevertheless a tendency for all prices to move up and down as a group. This is what is meant by refer-

ences to changes in the general price level. Throughout the world's history there has been a strong tendency for prices to increase — in spite of temporary setbacks. It is estimated that prices today are on the average from five to ten times as high as they were a thousand years ago. A pronounced upward movement of prices took place between 1500 and 1650, beginning in Spain, spreading then to other countries of western Europe. In terms of silver, Spanish prices more than tripled during the sixteenth century.¹ A similar peak was reached in England by 1650. In the United States the price level fluctuated moderately during the generation before the Civil War. From 1860 to 1865, prices as a whole more than doubled. Then there followed a steady decline till about 1897. During the next generation there was an almost uninterrupted rise, reaching its climax in 1920, when prices on the average were more than three times as high as in 1897. This climax was followed by a drop in 1921; but from 1922 to 1929 prices remained on a fairly stable level, about twice as high as in 1897 and about fifty per cent higher than in 1913. The stock market crash of 1929 was the noisy announcement of one of the most abrupt downward movements of prices in the world's history. By 1932 wholesale commodity prices reached a level below that of 1913, and they continued below the pre-War level until September, 1933. At the present writing commodity prices are about thirteen per cent higher than in 1913 and about twenty per cent lower than in 1926.

Complaints about prices. The importance which people justly attach to price is borne out by the frequent complaints which we hear about prices. During the first twenty years of the present century there was a steady stream of protest about the high prices of certain necessities of life, particularly the prices of meat and other foodstuffs, of coal and of house rents. People harked back to "the good old days" when butter was ten cents a pound, eggs fifteen cents a dozen, and coal five dollars a ton. Once the grim hand of depression was laid upon us, outcries arose of an entirely different nature. Business men complained because of the lowering of prices. Price-cutting was described as ruinous, and the man who lowered

¹ Earl J. Hamilton, *American Treasure and the Price Revolution in Spain, 1501-1650*, 1934, pp. 206-210.

his price was termed a "chiseler." But now that prices are rising again, the old demon, "high cost of living," stalks about promoting more unpleasantness. This is nothing new. There have been popular complaints against the high prices of the necessities of life ever since the Middle Ages, in fact ever since price began to be a matter of importance in the daily lives of the common people.

Business men in general are inclined to like high prices, because they appear at least to make business active and profitable. Individual business men however often complain that while prices in general are high, the prices of their particular products are unreasonably low. Such was the complaint of American farmers from 1923 to 1929, and later. In the face of the relatively high level of other prices during that period, wheat and other agricultural commodities were low, and this situation aroused bitter resentment among the farmers. This also was nothing new. The farmers of the United States have for generations been inclined to think that something was wrong with agricultural prices, and they have demanded that Congress or the state legislatures do something to relieve them. An understanding of current legislation and the political history of the United States since the Civil War is impossible without full recognition of this attitude.

Wage rates (the price of labor) appear to many to be unduly low. On the other hand, there are the big men in industry, railroad presidents, bank presidents, managers of great industrial enterprises, etc., whose salaries (the price of labor again), ranging to heights of \$50,000, \$75,000, or even more per year, sometimes appear to be too high, out of proportion to the value of the services rendered. Likewise the salaries of motion picture actors are sometimes objects of criticism, particularly when the remuneration of ordinary human activity declines. At certain times, as in the depression of 1938 and 1939, some people have believed that wage rates were too high.

Prices fixed by statute or custom. There was a time when prices were largely determined by custom and law. People in the Middle Ages talked of "just prices," "fair prices," etc. In general a price was just if it was as it always had been; the just price was the customary price. When a seller sought to exact a price higher than

what was customary he laid himself open to suspicion. So also of the buyer who sought to compel a price lower than the customary one. A price different from that established by custom was *prima facie* unjust. Custom was very powerful in those days, and the mere force of custom alone had great influence in preventing serious fluctuations in prices. To the aid of custom came law, civil and religious, and throughout the Middle Ages the law undertook, more or less completely, to prescribe what were the fair and just customary prices and to forbid departure therefrom. This situation has a certain parallel in modern times in the attitude of certain business men and labor unions which argue that prices and wages should not be altered because the alteration would involve a departure from "what is customary." But neither law nor custom was able permanently to prevent the price changes which followed in the wake of changed economic conditions. There was some attempt to adapt the law to these changes, and legal prices were altered from time to time, but the law was unable either to resist the changes or to adapt itself to them. Prices tended more and more to depart from the customary and legal schedules, some higher, some lower. Evidently there were forces at work more powerful than custom and human law.

The legal determination of price gradually declined and finally ceased to exert an important influence, though the policy has never completely disappeared. Some prices have always been subject to legal determination, even down to the present day, and movements to extend the sway of the law have appeared from time to time in the history of all countries. During the World War, Congress provided for the control of the prices of a great number of important commodities as a war measure. In the United States today legislative authority regulates the rates charged by the railroads for transporting both passengers and freight. Trolley and bus fares, telephone rates, telegraph rates, etc., are subject to legal control. Rates charged by gas companies, electric companies, water companies, and other local public utilities are generally determined or at least limited by legal enactments. But the mention of these cases is sufficient to remind us that they are the exceptions and not the rule and that prices in general are evidently not determined by legislation.

Custom likewise still plays some part in the control of price. The fees charged by physicians, lawyers, and other professional men are quite generally restricted by custom. Custom has an influence on the prices charged by barbers, bootblacks, cab drivers, etc. But it is evident that, in spite of custom, the prices even of these services do change and that there is at no time universal conformity to the customary schedule. What custom does in these cases is not to determine prices, but rather to act as a steadying force, keeping most prices in line with the customary standard and preventing many great and sudden changes. And after all, these cases again are the exceptions, not the rule. Prices in general are not determined by custom today.

Are sellers or buyers free to fix prices? People sometimes talk as though they thought the retail sellers of goods had complete power over prices. Does not the merchant, we are asked, fix his prices as he sees fit? What does the customer on the other side of the counter have to say about it? He can take the goods or leave them; that is all. But obviously things are not so simple as this for the retail merchant. In the first place, he was a buyer before he became a seller, and the prices he charges are obviously related to the prices he had to pay for the goods. Finally the buyer does have an influence on prices, exercised through his power to refuse to buy. The seller cannot put his prices too high; otherwise all or part of the goods will be left on his hands. On the other hand, the farmers often complain that in their markets it is the buyers who fix the prices of wheat and cotton, cattle and hogs. But as we shall later learn, even these buyers are limited by forces on the sellers' side over which the buyers have little or no control.

Is price a mystery? How is it then that prices come to be what they are? We have seen the extraordinary variety of prices. Some at least appear to be entirely freakish. A diamond may sell for \$1,000, a loaf of bread for ten cents. How is it that the mere luxury is worth ten thousand times as much as a prime necessity of life? A dress suit, worn only on rare occasions, may cost a hundred dollars; a suit for everyday wear, only forty-five. Why does the price of the less necessary thing exceed that of the more necessary? We have seen that prices are constantly changing, some rising while

others are falling, and the average of all prices moving steadily downward for years and then as steadily rising. Is all this complexity a mystery incapable of explanation? Is price merely a matter of chance? Or is it the natural result of causes which may be discovered? Certainly price is of enough importance in human affairs to make the search for its explanation worth while if there is promise that the inquiry will be rewarded by success. The investigation of the forces that determine price is the task next before us.

Consumer's choices. From the broad social point of view, goods are produced in order that human wants may be satisfied. From the viewpoint of the producer, he is engaged in production in order that he may sell his product to consumers for money, if possible for more money than the product cost him. The consumer finds himself confronted with an apparently infinite variety and quantity of goods more or less capable of satisfying his personal wants. He is invited to make his choice, to take whatever things he wishes and as much of each as he may desire, subject only to the condition that he give money in exchange in accordance with a stated scale of prices. This is indeed an onerous condition, and to the ordinary person it immediately imposes a stringent limitation upon his purchases. He abandons at once any idea he might have entertained of possessing the magnificent steam yacht; the price exceeds his total stock of money. He turns away from the tailor shop where a suit would be made to his order for \$75, even though he may have that sum of money or more. He finally makes his purchases — from a limited number of goods and in quantities of each which are strictly limited. His purchases are the result of a series of decisions, in the making of which he has had to weigh the relative merits of many competing wants, involving the exercise of more or less careful thought and judgment. It is the sum total of these individual judgments and choices of consumers that makes up the demand for goods.

Although it is true that the total demand of consumers is one of the chief price-determining forces, to any individual the prices of the goods in which he is interested appear as fixed quantities over which he has no control. It is estimated that in 1929 the people spent in retail markets of the United States almost \$50,000,000,000 ;

and in 1933, somewhat more than half that amount. The annual purchases of any ordinary person were obviously only an insignificant fraction of either of these figures. No one person spent a significant fraction of the 447 million dollars spent for radio sets, parts, and accessories in the year 1927 or of the $5\frac{1}{2}$ billion dollars spent in grocery stores and retail meat markets in the year 1933. Since therefore the ordinary individual's purchases are generally only a minute fraction of the total quantity of any good offered for sale, we are justified for the purposes of the present inquiry in assuming, as does the purchaser himself, that existing prices are independent quantities over which he has no control.

Everyone is vitally interested, as a consumer, in the problem of so disposing his limited stock of money among the unlimited number and quantity of goods offered to his choice that his total enjoyment may be as great as possible. And since it is these choices of consumers which determine the kind and the quantities of the things that may be sold, every producer is vitally interested in the mental processes by which consumers arrive at their choices. The exploration of this territory is the first step in the search for the laws of price.

Marginal utility: The subjective side of utility. Suppose you are playing golf on a day when the clubhouse employees are away and it is impossible to buy golf balls at the club. Your drive off the first tee goes into the rough, and the ball is lost. If it happens to be the only golf ball in your possession, your chagrin at its loss will be extreme, far greater than is measured by the price paid for that particular ball. Seeing your whole day's sport gone, you say, "I would give five dollars for a golf ball right now." Five dollars then may be taken as the measure of your loss. Now let us change the supposition by assuming that your bag is packed with a dozen new balls. The loss of the first drive now means probably nothing more than the money value of the ball. If, instead of a dozen, you had had half a dozen on which to rely for the day, the loss of one would have been a more serious matter. If you had started with only four, or three, or two, the loss would have been more serious still. Changing the illustration again, let us suppose that as you play through the course you meet a thrifty caddy who

offers to sell a golf ball which he has found, but only at the highest possible price for which he can hold you up. Ignoring any restraining club rule, how do you decide what you will pay? The first thing is certainly to take count of your stock. If you happen to have just driven your last one into the woods, you will if necessary pay five dollars or whatever sum measures to you the satisfaction of continuing the game. If you have just one golf ball left, your offer will not be so high, but it will still be high. If you have two, or three, or four, your offer will be correspondingly lower, till finally you may be so well supplied that you will decline to pay more than the regular price or even refuse to buy at all. We can readily draw the general conclusion that the utility of an additional golf ball depends at any given time upon the number possessed, varying inversely with the number possessed, and that the same rule holds of the utility of a golf ball subtracted from the stock on hand.

This example will serve to introduce a human characteristic which relates itself to all goods (wealth and services). Heretofore we have studied utility objectively, as that quality of wealth and of free persons which enables them to satisfy human wants. Here we encounter utility subjectively, from the viewpoint of the person whose want may be satisfied. Objectively a useful thing "satisfies a human want;" subjectively the person "realizes satisfaction" from the useful thing when it satisfies his want. *Utility thus expresses the satisfaction which a person expects to realize from a commodity or service.*

When a person apportions a given quantity of any good to different uses, the wants to be satisfied by the several units are not generally of equal importance. Suppose there is a water shortage and a family finds itself in possession of ten buckets of water for the day's use. One bucket might be set aside for drinking, one for cooking, one or two for washing the dishes, two or three for bathing, and so on down the list of uses of water in a descending scale of importance. If the number of buckets on hand were less, less water would be devoted to each use. But it would be the less important uses that would suffer the greater curtailment; some one or more of the less important uses might have to be given up entirely; baths might be omitted. If the quantity were greater,

other uses still less important might appear. In this example the wants to be satisfied by the several units are different in kind as well as in importance, but the principle holds equally when the wants are all of the same kind, as in our example of the golf player.

The nature of utility. Strictly speaking, utility is something that cannot be measured. Attempts to measure utility in terms of money break down because the amount of money one would give for a unit of any good is determined, not only by the satisfaction anticipated from the use of the good in question, but also by the satisfactions to be derived from other goods which might be purchased with the same sum. Our golf player was under certain conditions willing to pay \$5.00 for a golf ball, because he anticipated more satisfaction from its use than he could derive from any of the innumerable other goods which \$5.00 would buy. Obviously the quantities of other things \$5.00 would buy depend upon the prices of the other goods. Again the money one would pay for anything depends upon the amount of money one has. A rich man may derive no more satisfaction from playing golf than a poor man, but given the desire to play, he would be willing to pay more for a golf ball.

In fact the only way we can tell what is the utility of a good to an individual is by observing the way in which the individual will choose between the purchase of the good in question and some other way of spending his money. In other words, when we speak of the utility of any good to an individual, we are really trying to describe the way in which he will choose between two or more alternatives. When we study the behavior of a consumer, we are examining the way in which he will distribute his income among different lines of expenditure. Nevertheless it is possible, if not to measure utility itself, at least to measure the relative utilities of two objects to a given person. Reverting to our former illustration, let us suppose that you happen to have with you five packages of chewing gum when the caddy appears and offers to exchange a golf ball for a package of gum. If you have started out very short of golf balls, it is probable that you will make the exchange, even though the golf ball may already be well worn and conceivably worth less than five cents. If now the caddy were to offer you a second golf ball on

the same terms, you would view the proposed transaction with more doubt. Your situation would have been changed in two respects by the first transaction. In the first place, you would have one more golf ball than you began with; in the second place, you would have one less package of gum. You might now offer the caddy four sticks for the second golf ball but not an entire package. Similarly with a third exchange. By that time the chances of your needing another golf ball would be even smaller, whereas the importance of each stick of gum would have increased. You might carry through a third exchange if the caddy would part with a golf ball for two sticks of gum; almost certainly you would not make a third exchange on the same terms as those on which you made the first or the second.

Now it is important to realize that no friend of yours could have said what was the utility of a single golf ball to you, either when you went out on the course or after the first or second exchange. Nor would you yourself be able to attach any absolute value to or in any way to measure the utility either of a golf ball or of a single stick of gum at any point in the story. For this reason we cannot with confidence say that the utility of one added golf ball to you would be less after the second exchange than after the first. But what anyone present at the transaction could tell from your behavior is that gum was becoming relatively more valuable to you than golf balls as these exchanges proceeded, and it was probably true also that golf balls were becoming relatively more valuable than additional sticks of gum to the caddy. To sum up, it is quite possible to compare accurately the utilities of two different objects to an individual even though it is impossible to measure utility itself.

Relative marginal utility. This fact will be useful to us in our analysis of the behavior of consumers, since every purchase of an object may be thought of as an exchange of money for the good in question. Now to the individual money is ordinarily not a thing of intrinsic worth. It merely represents a certain quantity of purchasing power, which may be used to buy any of the things that normally enter into his consumption. Therefore the purchase of, let us say, a golf ball is really a transaction in which the individual

exchanges for the golf ball a certain amount of purchasing power, which is valuable because of all the other things which it could have bought. In the example we have just discussed we can say that at the beginning of the match, before any exchanges had taken place, one more golf ball would just compensate the player for the sacrifice of five sticks of gum. After the first exchange one golf ball would just compensate him for the sacrifice of four sticks of gum, and so on.

Let us now refer back to the player's original statement quoted above, "I would give \$5.00 for a golf ball right now." What this may be said to mean is that one more golf ball at that particular moment would just compensate him for the sacrifice of \$5.00; that is, for \$5.00 worth of all the other things that he might purchase. This, as the reader will recall, was when his stock was zero. If instead he had one golf ball, another would not in his estimation be as good as \$5.00 worth of other things, but something less, say \$4.00. Using figures thus which, though arbitrary as to absolute amount, are in the proper relation to one another, we may set up a schedule showing the amount of money which he would be willing to sacrifice (if he had to) in order to change his stock of golf balls by one unit under various numerical circumstances:

<i>Quantity possessed</i>	<i>Number of cents which would be sacrificed in order to gain one unit</i>	<i>Number of cents which would just compensate for the loss of one unit</i>
0	500	—
1	400	500
2	315	400
3	240	315
4	185	240
5	140	185
6	105	140
7	75	105
8	50	75
9	30	50
10	15	30
11	5	15
12	0	5
13	-5	0
14	-7	-5

This schedule represents the possibility that a person may have so much of a commodity that there is no sum of money so small that an added unit (a thirteenth golf ball) would compensate for its loss. In other words, he will not pay so much as a cent for one more unit. Beyond a certain point additional units of the commodity may indeed lessen his total satisfaction by burdening him or inconveniencing him to a certain degree. Thus extending the schedule with negative figures, we show the sums of money which he would have to be paid in order to induce him to take additional units.

Definition of marginal utility. Either the second or the third column in the preceding table may be logically headed "ratio of the marginal utility of golf balls to that of money" in accordance with the following general definition: *The ratio of the marginal utility of any good to that of another good for any person is equal to the number of units of the first good that would just compensate him for a change of one unit in the quantity of the second good.* For most purposes it is not convenient to speak of the ratio of the marginal utilities of two different commodities, and realistically in the modern world individual consumers do not think in terms of barter, because the typical transaction which any consumer carries through is one in which he exchanges a certain sum of money for a good or a service. It is usually convenient therefore to measure the tastes of a consumer in terms of *the ratio of the marginal utility of some particular good or service to that of money.* This ratio we shall define as the *relative marginal utility of the good* or, for the sake of brevity, as its *marginal utility*. Hereafter the term marginal utility may be understood to mean "relative marginal utility."

In the application of this concept however a certain convenient numerical simplification is introduced to avoid confusion. "A change of one unit" may occur either through addition or through subtraction. Reference is more frequently made to the "addition of a unit"; but the "loss of a unit," or the "subtraction of a unit," is an equally important possibility. In order not to have two ratios of marginal utilities for each and every quantity, the marginal utility is generally understood to be the marginal utility of a quantity *which includes the unit added.* Thus relative marginal utility through addition and through subtraction is gen-

erally considered the same for any particular quantity. Revising the preceding table in accordance with this convention, we have a marginal utility schedule as follows :

<i>Quantity</i>	<i>Marginal utility</i>
1	500
2	400
3	315
4	240
5	185
6	140
7	105
8	75
9	50
10	30
11	15
12	5
13	0
14	-5

Graphical representation : A digression. Study of the facts presented in a schedule is generally made easier if the schedule is translated into a curve or graph. Such graphical aids to the presentation and interpretation of statistical material have proved themselves very useful both in the study of economics and in the conduct of practical business affairs. For this reason and because we shall have frequent occasion to use the graphical method of presentation in this book, we may well take this opportunity to inquire into the essential features of the most commonly used type of graph.

Referring to the utility schedule above, we note that it presents two separate magnitudes. The problem is to construct a graph which will show the relations between these magnitudes. We first provide a chart, prepared in the following manner (see Figure 1). Draw a horizontal line parallel to the bottom of the page. This line is technically known as "the axis of abscissas," but for present purposes it may be more simply designated as the *horizontal axis*. Toward the left-hand edge of the page draw a vertical line, intersecting the horizontal axis at right angles. This vertical line is technically known as "the axis of ordinates," but we will call it the *vertical axis*. The point where the two axes intersect, in the lower left-hand corner of the chart, is called the *origin*. This point, marked *O*, represents zero on two scales, a horizontal scale and a

vertical scale. It is customary to mark the right-hand extremity of the horizontal axis x ; and the upper extremity of the vertical axis y . The horizontal axis may thus be referred to as the x axis, or Ox ; and the vertical axis, as the y axis, or Oy .

On the horizontal axis mark off a scale to represent quantities; *i.e.*, in this particular case numbers of golf balls. Begin this scale with zero at the origin, and mark it off evenly along the axis with appropriate numbers. The axis is thus divided into a number of

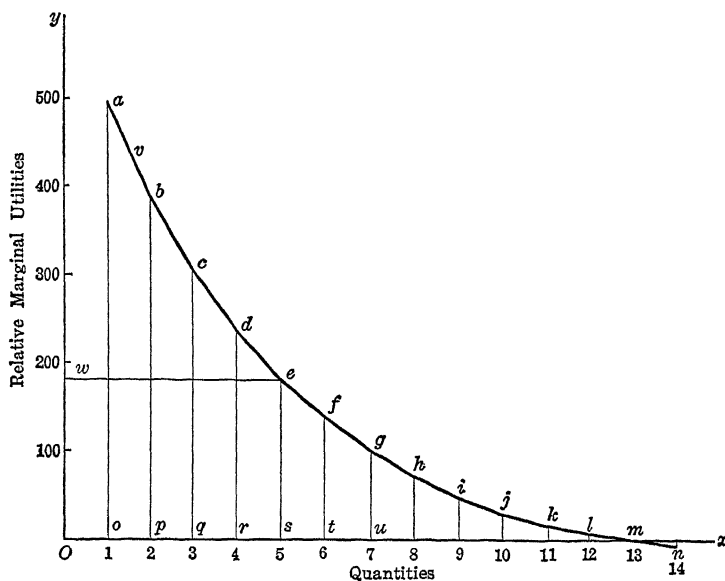


FIG. 1

equal parts called scale divisions. The greatest of the numbers on the scale should be at least as large as the largest quantity to be represented in the chart. In this example the largest quantity is 14. Spread the scale so that the largest given quantity appears well over to the right; this is in order that the magnitudes we deal with may not be crowded too closely into the left-hand corner of the chart. Skip no intervening numbers between zero and higher magnitudes; let this scale be continuous from left to right.

In almost all English and American economic charts, quantities of goods are measured along the horizontal axis; and other magni-

tudes (utilities, prices, costs of production, etc.) along the vertical axis. This is a general custom, purely arbitrary, but very convenient. It avoids confusion. To make the chart absolutely clear however it is advisable to print the word "Quantities" along the horizontal axis, and another appropriate word or phrase along the vertical axis.

In this case we print the words "Relative marginal utilities" along the vertical axis, and along this axis we mark off a scale. Begin this scale, as well as the other one, with zero at the origin. Mark it off evenly upward with appropriate numbers. Let the greatest of these numbers be at least as large as the largest measure of utility to be represented in the chart. In this case this largest measure is 500. Spread the scale so that the largest given measure of utility appears roughly as far above the origin as the largest given quantity does to the right of the origin. The scale divisions on either of the two axes need not, and usually will not, be equal to those of the other. In the vertical scale, as in the horizontal one, skip no intervening numbers between zero and higher magnitudes. If these initial directions are followed, the entire figure when completed will comfortably cover the designated space; the scales can be easily and correctly read; and the inclined lines in the figure will not be too steep or too nearly horizontal for quick, accurate interpretation.

Now let us proceed to locate on the chart points representing the respective relations between the possible number of golf balls, from one to fourteen, and corresponding relative marginal utilities. The first line of the utility schedule indicates that the marginal utility of one golf ball is 500. To represent this supposition graphically, locate 1 on the horizontal scale (as shown in Figure 1), and proceed in a straight line vertically upward until a point is reached exactly on a level with 500 on the vertical scale; or from 500 on the vertical scale proceed in a straight horizontal line to the right until a point is reached exactly above 1 on the horizontal scale. In either case we arrive at the point marked *a* in Figure 1. This point represents the supposed fact that the marginal utility of one golf ball is 500.

According to the second line of the utility schedule, the marginal utility of two golf balls is 400. Locate 2 on the horizontal scale, and proceed in a straight line vertically upward until a point is reached

exactly on a level with 400 on the vertical scale; or from 400 on the vertical scale proceed in a straight horizontal line to the right until a point is reached exactly above 2 on the horizontal scale. By either method one arrives at the point marked *b* in Figure 1.

In similar fashion twelve other points (*c, d, . . . n*) may be located, representing twelve other suppositions regarding the marginal utilities of various numbers of golf balls as shown in the utility schedule. Fourteen points in all are located. They are connected by a series of lines. Thus a "curve" *an* is obtained, which is known as a *relative marginal utility curve*.

Having constructed the curve, we pause to consider its interpretation. It is a marginal utility schedule, a golf player's mental attitude, converted into graphical form. Every significant point on the curve bears a definite relationship both to the horizontal and to the vertical axis. The distance of such a point from the vertical axis, as measured by the horizontal scale, indicates a certain number of golf balls. The height of this point above the horizontal axis, as measured by the vertical scale, indicates the amount of money our imaginary player would be willing to sacrifice to obtain a golf ball which increases his total stock to the number indicated.

Certain additional lines in Figure 1 would not ordinarily be necessary, but they have been drawn to assist us in detailed study of the chart. Consider, for example, the line *we*, drawn from *e* perpendicular to *Oy*. Its length, the same as *Os* on the horizontal axis, indicates that the point *e* refers to 5 golf balls. In conformity with this information the height of the point *e*, equal to *se* or *Ow*, indicates that if a fifth golf ball is acquired by the player, it will just compensate him for the loss of 185 cents. The height of the point *e* likewise indicates that the loss of any one of a total of five golf balls could be compensated by the sum of 185 cents. From points *a, b, c, d, f, . . . n*, other similar definite readings may be made.

Points on the curve *an* (Figure 1) other than those marked by the letters, *a, b, c, . . . n*, have no significance. Since we cannot deal with fractions of a golf ball, an intermediate point, as *v*, means nothing. Strictly the diagram should not contain a curve at all, but merely a series of dots, or points. The points are joined by a curve simply in order to bring out more clearly the relation between

marginal utility and the quantity of the good under consideration. When it is possible to use very small units, the points may become so numerous as to give us practically a smooth curve. Thus if we represented the marginal utility of sugar, measured in pounds, the diagram would be somewhat as in Figure 1. Measured in quarter-pounds, we should have four times as many significant points. Measured in grains, the points would be so close to each other as to form practically a continuous curve.

Actually we seldom are able to use units small enough for this result, but, when dealing with units which are capable of subdivi-

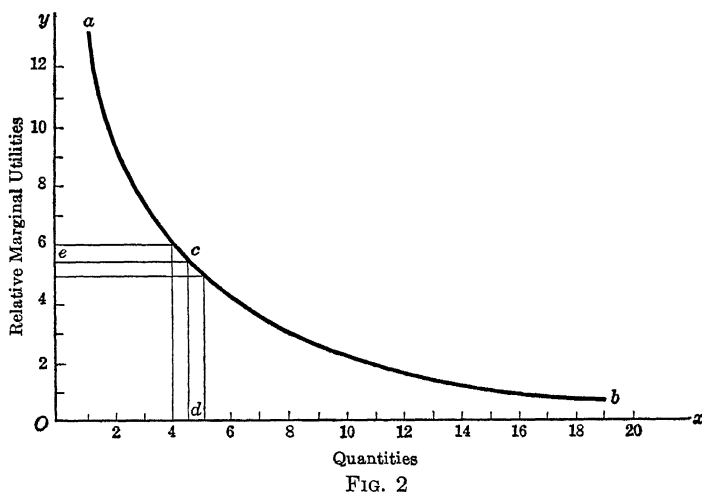


FIG. 2

sion, it is possible to carry the graph to a further stage of usefulness. For example, we may suppose that the marginal utility of sugar to a certain person were as shown in the above diagram (Figure 2). The schedule from which this curve was plotted presumably was expressed in terms of pounds and cents. Let us suppose now it is desired to know the marginal utility of $4\frac{1}{2}$ pounds of sugar. Obviously this will be somewhere between 6, corresponding to 4 pounds, and 5, corresponding to 5 pounds. To obtain the answer graphically, it is first necessary to change the curve from a series of straight lines connecting the ends of the several ordinates to a curved line passing through them. This "smoothing of the curve" has been done in Figure 2. The marginal utility corresponding to

$4\frac{1}{2}$ pounds is now readily shown by the point *c*, indicating that when $4\frac{1}{2}$ pounds (*Od*) are considered the marginal utility is $5\frac{1}{2}$ (*Oe*). This process is called *interpolation*.

The simplicity of the illustration used in this elementary discussion of the graphical method does not adequately emphasize the services of the curve as an aid in analysis of statistical data. The student of intricate economic phenomena in the business world finds that so vast and complex is the mass of data with which he must deal that other methods of analysis are inadequate. Of course the curve adds nothing to the evidence of the data from which it is constructed; it does however reduce complex data to a simplified form, so that inherent truths may be brought to light, thus abbreviating the search for the laws which govern the relations of different factors. Even where other methods are effective, the graphical method is usually serviceable for preliminary and supplementary studies. It will be called into frequent service in the following pages of this book.

Meaning of marginal utility. It must be emphasized that our marginal utility schedules and curves measure relative marginal utilities. Because utility cannot be measured and because all that we know either about ourselves or about others is the way in which we will choose between alternatives that are presented to us, the concept of an absolute marginal utility would have no meaning.

Since the marginal utility of a good, as we define it, is relative to the marginal utility of money, its marginal utility will change with changes in the amount of money possessed even when the amount of the good itself remains the same. If for instance you were suddenly given a thousand dollars to spend within the next ten days, the relative marginal utility of practically everything that enters into your budget would be sharply increased. The income you would have available for expenditure during the next ten days would suddenly have become so abundant that the sacrifice of five dollars for a golf ball would seem much less important than before. The additional satisfaction you would gain by receiving a golf ball after you had lost your only one at the beginning of a round would be much greater than before measured in terms of money; that is, in terms of the other things you might buy with any given number

of dollars. Similarly the relative marginal utility of any good to a wealthy man will usually be much greater than to a poor man unless the former already possesses far more of the good in question than the latter.

We must also remember that a marginal utility curve does not show a historical sequence of uses or utilities ; it depicts a series of simultaneous possibilities — not in the sense that all of them can be realized at the same time, but in the sense that any of them may be realized under the same general circumstances. Thus at a given moment our imaginary player may have a dozen golf balls, or he may have none at all, or he may have any intermediate number. The relative marginal utility curve shows the additional satisfaction in terms of money he derives from any golf ball which increases his total stock to any indicated number within a given range of possibilities.

It is somewhat misleading to say, as is sometimes done, that “marginal utility is the utility of the last unit.” The units are all alike. Once in the possession of the player any unit will render him equal service, afford him equal satisfaction, no matter in what order it may have been acquired. If a difference in utility is occasioned by the addition or subtraction of a particular unit, this change cannot be attributed to any peculiar characteristic of that unit. Any other unit would produce the same result.

The law of marginal utility. Once we have a clear idea of the meaning of relative marginal utility, the economic law of the relation between the quantity of an economic good and its marginal utility becomes obvious. This principle is known as the *law of marginal utility* and may be stated as follows: *At any given time the relative marginal utility of any good to any person varies inversely with the quantity under consideration.* A glance at Figures 1 and 2 will make the meaning of this perfectly clear.

This law merely expresses a psychological fact with which we are all familiar. The appetite for food is the simplest illustration. Having eaten one apple, the second gives us less satisfaction, the third still less, and very soon we reach the point of satiety, when no further satisfaction, perhaps even repugnance, results. The same loss of appetite would have appeared more gradually if we had

considered the separate bites. This principle will be recognized at once as holding good of the satisfaction of practically all wants, though there may appear to be exceptions when for a time the satisfaction seems to increase, as, for example, when one listens to music and, gradually getting into the spirit of it, gets more pleasure from the second piece than from the first. Such cases are not true exceptions since the point of diminishing utility is sure to be reached sooner or later, which is the essential feature of the law of marginal utility.

It should be realized however that the sort of marginal utility which commonly determines economic action does not arise from immediate use or consumption, but from anticipation. When the housewife buys sugar, she does not generally use it and pay for it pound by pound. Ordinarily she buys enough to last at least several days. For purposes of immediate consumption the marginal utility to her of a relatively small amount of sugar is zero; but for the anticipated uses of the next week or ten days the same amount of sugar may have a high marginal utility. Therefore she buys not only this amount but also several pounds more. To the psychological principle of satiety must therefore be added the principle that man takes account of the future and makes provision for his needs beyond the present moment. But just as there is a satiety in immediate consumption, there is a like principle in anticipation of future satisfactions. The greater the amount bought for future use, the less is the added prospect of satisfaction through an added unit of the commodity. Hence the law of marginal utility applies, not only to one's experiences of the present, but also to one's outlook on the future.

Marginal utility of durable goods. The principle of marginal utility has been developed thus far with respect to commodities which are rather quickly consumed and of which a person therefore usually possesses several units. On the other hand, one does not usually consider buying more than one lawn-mower, or several pianos or dwelling houses. These are durable goods, each furnishing a series of uses for a comparatively long time. One does not ordinarily need more than one unit of such a good. How can the principle of marginal utility apply to such a commodity? In answer we

should note first that it is the series of satisfactions from the use of a durable commodity that furnishes the scale of diminishing marginal utility. When a person is faced with the alternative of buying one piano or none, he considers the series of uses which a piano would promise him through the years to come. The relative marginal utility of the whole series, in connection with the price of the piano determines his decision. Again we note that, having the choice between different qualities of durable goods, a person is often able to fix upon a certain quantity of services in accordance with the principle of marginal utility. In deciding whether to buy a house for \$2,000 or \$20,000, one is determining how much "house service" he will buy. The relative marginal utility of a unit of service of the \$2,000 house — furnishing not much more than the bare requirements of shelter — will be far greater than that of the much greater series of services — including spaciousness, artistic satisfaction, and the gratification of vanity — which the more costly house offers. Even the durable goods, of which only one unit is generally possessed, thus conform to the general principle of marginal utility.

Some apparent price anomalies resolved. In the concept of relative marginal utility lies the key to certain price relations which sometimes appear to be paradoxical. How is it that some of the most useful things in the world are so cheap, while other things that could be given up without any real suffering are so costly? To some the fact that a diamond may be worth a thousand dollars while a loaf of bread is worth ten cents seems strange, inexplicable, or even wrong. The answer is that the value of anything depends, not upon somebody's judgment of the importance or righteousness of the want satisfied, but on demand and supply, and the demand side is controlled by considerations of marginal utility. Marginal utility, we have just seen, varies inversely with the quantity. Diamonds are scarce; their marginal utility is therefore high. Bread is very abundant; its marginal utility is correspondingly low. This means that people are willing to pay high prices for diamonds, but very little for a loaf of bread.

Abundance and scarcity. Some of the most useful things in the world are so abundant that they have no marginal utility at all and

therefore no value. Among the prime necessities which man must have in order to sustain life is air ; nothing could be more useful. Yet no one will pay for air, simply because it is so abundant that everyone can have all he wants for nothing. Its marginal utility is zero. To have value, a thing must have not only utility, but a marginal utility greater than zero. The only things that can have positive marginal utility are those which both have utility and are so limited in quantity that there is not enough to satisfy everybody's wants. It is for this reason that some of the most important things in the world, as sunlight, air, sometimes water, are not included in the subject matter of economics ; they are not economic goods. They have no value, are not bought and sold ; it is this, rather than any permanent physical characteristic, that rules them out. Water, free in abundance for all in many country districts, is limited in the city ; it there has a significant marginal utility, has value, and is an economic good. Land in certain newly settled regions is sometimes so abundant as to be valueless. This condition is almost sure to pass, and in general land is so strictly limited and so insufficient to satisfy all human wants that its relative marginal utility and its value are usually high.

EXERCISES

1. The marginal utility schedule below reflects the subjective estimate of an individual of the importance to him of successive additional pairs of shoes :

<i>Quantity</i>	<i>Marginal utility</i>
1	\$100
2	20
3	12
4	6
5	2
6	1

- (a) Following the instructions contained in this chapter, construct a diagram to represent this marginal utility schedule.
 - (b) If the particular grade of shoes sells for \$10 a pair, how many pairs will this individual find it worth-while to possess ? Explain.
2. A college student has \$3 to spend for a ticket to a football game or a dance or a book or a new hat. Explain the basis on which he makes his decision between these alternative forms of expenditure.
3. Assuming that people are frequently influenced by shrewd salesmen, misleading advertising, gossip of friends, etc., to buy goods which give them little real satisfaction, is this inconsistent with the principle of marginal utility ?

[illegible]

In the early days of the science of economics, a cynical observer is reported to have said: "If you want to make a first class economist, catch a parrot and teach him to say 'supply and demand' in response to every question you ask him." It must be admitted that there is a common impression that the whole solution of the problem of price is contained in the statement that "price is determined by supply and demand." It is true that the prices of many goods are thus determined, but the problem of price is not quite so simple as this statement would imply or as is commonly assumed.

This has already been illustrated by the case of an imaginary golf player. The curve *an*, in Figure 3, represents the conditions assumed in that example. It shows the money estimation of the loss which a certain individual would suffer by parting with one golf ball, assuming him to have any given number on hand, or the price he would be willing to pay for one additional golf ball. Every vertical distance (such as *ao*) thus measures the price one would be willing to pay for the additional unit which would bring his stock up to the number indicated on the base line, provided of course he were held up each time for his highest price.

Now this of course is not the way golf balls, or other things, are ordinarily bought. The somewhat frivolous example which served us in developing the notion of marginal utility may now be discarded in favor of the more normal state of affairs. There is an established price on the market. No matter how badly the buyer may want the first few units, no matter how great their utility to him, no matter how much he might be willing to give for them if he had to, he does not have to pay more than the market price; and he

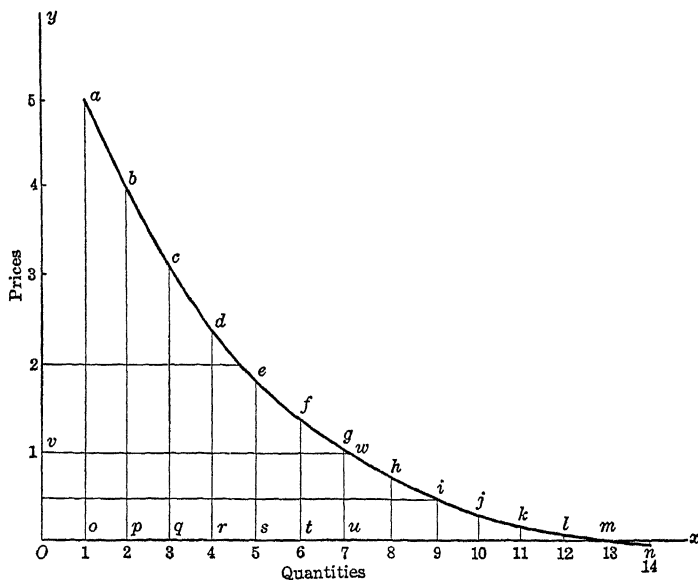


FIG. 3

may ordinarily buy all he desires at that price. The question he has to determine is, not what price he will pay, but how many he will take at the established price. Suppose that our sportsman, laying in a stock of golf balls for the week-end, finds that the price of the particular make he uses is one dollar. How many will he buy? He will certainly buy one; rather than go without one he would be willing to pay five dollars. Similarly he would buy a second, a third, a fourth, a fifth, a sixth, and a seventh since in each case the sum he would give if he had to is greater than the price. An eighth ball is worth to him only seventy-five cents. He would gladly have this

one and several more, but not at a dollar. We conclude that this person would buy seven golf balls.

The curve in Figure 3, which originally illustrated the relative marginal utility of a golf ball under different circumstances, will serve equally to illustrate the quantities that this individual would buy at various prices. From the point on the y axis which measures one dollar, we draw a horizontal line, vw . This line intersects the curve just below the point g , indicating that at a price of one dollar seven golf balls will be bought. Again suppose our golf player had found that they were charging two dollars. He would consider this an exorbitant price and, while he would have to get enough to carry him through the week-end, he would buy fewer than if the price were more reasonable; in this case four. If, on the other hand, he had found that his favorite brand was being sold at fifty cents, a bargain price, he would have bought a few more; namely, nine.

Of course only one of these price possibilities can be realized at a given time. Suppose the price is actually announced to be one dollar. The individual will buy seven units. This act should not blind us to the fact that consciously or unconsciously there resides in him at the same time both the willingness and the ability to buy nine units at fifty cents, four units at two dollars, and other amounts at other prices as shown in the schedule. Such ability and willingness constitute an individual's demand for a commodity. The concept may be defined thus: *An individual's demand for any good (wealth or service) is a schedule of the respective quantities of that good which he is ready to buy at all possible prices.*

A curve such as that in Figure 3 is the graphical representation of this schedule and is known as the individual demand curve. It is identical with the relative marginal utility curve. Whichever name it takes, the curve is a picture of possible action in response to possible prices. It is controlled, not by the marginal desirability of the specified commodity alone, but by a balance of the desirability of that commodity against the desirability of other commodities and services.

Significance of this analysis. Hardly anyone would actually take the trouble to make minute calculations such as have been assumed in the foregoing analysis. No one could, because there is no such

thing as precise measurement of utility. Nevertheless most of us do have to exercise judgment in the expenditure of money, whether it be only a dollar or ten thousand dollars. Everyone has at least in his subconscious mind some idea of the satisfaction he expects to get from his purchases and can sense to some extent the result of spending a little more on one thing and correspondingly less on others. The given examples are more precise and definite than ordinary human behavior, but such examples are necessary for an insight into the complexities of human problems. A map is far simpler than the country it describes. A globe in the library is smoother than the earth it represents. For the same reason illustrations of economic principles often have to be simplified at the expense of absolute realism. The preceding analysis illustrates (admittedly in ultra-simple and definitive fashion) the manner in which an individual weighs marginal utilities against each other, adjusting his expenditures when he finds marginal utilities out of balance, seeking marginal utilities proportional to the prices he has to pay — all in the effort to make the most of the limited purchasing power at his command. The thought, the feeling, the unseen reflexes which govern this balancing of utilities, may be known only to the individual himself, and even then very imperfectly understood; but out of it all arises visible action, the purchase of economic goods. We are interested, not only in what the individual actually buys, but also in all that he is able and willing to buy under various possible circumstances.

How both parties to a trade can gain. People sometimes find it hard to believe that both parties to a trade can gain. It has been somewhat plausibly argued that, since in any honest trade the values of the two things exchanged must be the same, neither party can have made a gain; one party can gain only by giving less value than he receives; that is, by cheating the other party. This notion influenced the churchmen and lawmakers of the Middle Ages to believe that trade should be strictly regulated and customary prices enforced in order that neither party to a trade should gain at the expense of the other.

But unless the relative marginal utilities a buyer has in mind are born in ignorance and false anticipation, he gains by the oppor-

tunity to purchase any commodities so long as the amount obtainable for one dollar more than compensates him for the loss of the dollar. The seller, on the other hand, has received more than enough money to compensate him for the loss of the goods with which he parted. This side of the trade will be more clearly set forth in a later chapter. In the meantime, Figure 3 gives us an indication of the fact that the buyer has gained. Suppose that the price is one dollar. This individual would buy seven units. From the shape of his demand curve, which measures the relative marginal utility to him of various amounts of the good, we can tell that the first unit of the good would more than compensate him for the loss of five dollars, the second for the loss of four dollars, and so on. We can be certain that each of the seven units which he purchases more than compensates him for the dollar which he must spend to acquire it and that if his judgment has been sound he is indubitably better off after the transaction than he was before.

Things one cannot afford. Emphasis has been laid on the balancing of marginal utilities; *i.e.*, the adjustment of purchases in such a way that marginal utilities are proportional to prices. It is evident however that all the objects of an ordinary person's desires cannot be fitted into such a scheme of things. An average man does not have his clothes made to order in an exclusive tailor shop, even though his ambition may run toward sartorial elegance. Half of his monthly salary check might more than cover the bill; but, if the thought occurs to him, he dismisses it, saying, "I just can't afford it." Such remarks are common. They mean simply that spending money for a certain purpose, and thus having less for other wants and needs, would throw marginal utilities out of balance. With limited purchasing power, only a limited number of things can be bought, and only a limited number of marginal utilities can be balanced against each other. For the average individual many marginal utilities, though high, are not so high, even for the smallest quantities which can be bought, as other marginal utilities for somewhat larger quantities in proportion to corresponding prices. Thus many marginal utilities have to be left entirely out of the general balance.

Real people. Of course the ordinary person does not use marginal utility curves or schedules or demand curves or schedules in determining what he is actually going to do; they are for the economist ways of sketching the various things an individual may do in response to various possible stimuli. The closest approach to rational buying is seen when people budget all their expenditures. A careful family budget is an unconscious illustration of the principle of balancing marginal utilities. But many, perhaps most, people do not go even so far as the family budget. Poor people and even the moderately well-to-do often appear to spend their limited stores of money in the most foolish and irrational manner. A family that is without means to provide warm clothing for the coming winter will cheerfully spend ten dollars on a day's outing at the seashore. The radio, with its monthly installment payments, may be eating up the money that would seem much better spent for nourishing food. The daily attendance on the "movies" seems quite superior to any law of diminishing utility. Clever advertising and smart salesmanship bank on the careless and irrational way in which people spend their money. All sorts of appeals are used to induce the careless purchaser to buy things the utility of which seems absurdly small in comparison with the money expended. To some the appeal of a "bargain" is irresistible, and people come home laden with things for which they have no real need, bought simply because they were cheap; every retail merchant takes advantage of this human characteristic. In any city it is possible to see people buying things in one store when the identical articles are on sale at a lower price only a block or two away. A study of the purchasing habits of the general public is apt to induce a low regard for average human intelligence.

Is all buying irrational? From such facts as these the conclusion is sometimes drawn that there is no such thing as rational judgment in buying, that the principle of balancing utilities has no existence outside the imagination of the theoretical economist, and that all our talk of demand and supply is pure theory without any useful application to the hard facts of the practical world.

But let us not be too hasty in jumping at such conclusions. For

one thing, it will be well to recognize that much of the conduct of other people which we regard as irrational is not really so but is merely at variance with our own judgment. We readily think we could lay out a family budget for our impecunious neighbor which would materially increase the utility of his purchases. But it is his judgment — such as it is — not ours, which decides how he can get the most for his money.

The well-fed, well-dressed woman of the “upper class” points with scorn at the “irrational” conduct of the stenographer taking her pathetically inadequate luncheon at the soda fountain in order to save money to buy silk stockings. But this conduct, whatever else may be said of it, is not irrational. The girl purposely spends more money on clothes, and less on food; this is her way of balancing relative marginal utilities. Her conduct may be “foolish” according to some other person’s judgment. But the economist is investigating how she actually spends her money, not how she “ought” to spend it. He can only infer from the fact that she does spend her money on clothes that their relative marginal utility to her is higher than that of the food which she is thus prevented from buying.

Let us reverse the case. The shopgirl laughs at the wealthy patrons of the Fifth Avenue store who “haven’t sense enough” to walk over to Sixth Avenue and buy the same things at lower prices. But it does not follow that the wealthy shopper is irrational. She wants the calm restfulness, the air of luxury and dignity, the deferential personal attention of the exclusive Fifth Avenue shop. These are a part of what she is paying for. She knows she could save money on Sixth Avenue, but the loss of these other advantages would mean a net sacrifice in the utility of her purchases. She also may be “foolish” according to some other person’s standard, but that is beside the present point.

Habit and utility. Objection is sometimes made to marginal utility analysis on the basis of the old adage that “man is a bundle of habits.” Thus it is claimed that people buy things simply because they are accustomed to buying them, or because they see other people buying them and they want to “keep up with the Joneses.” There is truth in these statements, but they do not over-

throw the principle of balancing utilities. It must be recognized that utility is often nothing else than the capacity of a thing to satisfy a habit — the tobacco habit, for example, or the liquor habit, or if you wish a milder example, the Englishman's tea. We also see that the motive in the purchase of many things — a fur coat, a Rolls Royce, a costly mansion, or even a brilliant necktie — is often the desire to "show off" — the desire to impress people, to show that you are as good as they are and better. Professor Veblen has called this "conspicuous consumption." There are a great many elements governing human action, and reason is only one of them. Personal habits, social ambitions and social imitation, mass psychology and all that, obscure human reason; but they do not blot it out of existence. In customs and habits themselves there is generally some degree of reason. Furthermore it often takes a great deal of scheming to satisfy a habit, whether we regard it as "reasonable" or not. Witness the trouble people went to in order to get around prohibition. It often takes a great deal of scheming to keep up the same social display as other people in your own set. To accomplish it many other possible purchases may have to be sacrificed. Thus in habit and in social ambition there is still a balancing of marginal utilities. The economist does not begin by passing judgment on the various kinds and degrees of satisfaction people expect to get from their purchases, nor can he trace such things to their ultimate sources or ends. The economist asserts however that in general and at least in a rough sort of way each individual's expenditure of money is governed by a balancing of marginal utilities in accordance with the individual's own view of them in his own particular environment.

Total demand: Individual demand and total demand. Thus far we have considered the standpoint of the individual purchaser, who regards the prices of what he would buy as fixed and not subject to his control. But though the individual's demand for any good normally has no appreciable effect upon its price, the total demand of all consumers has a profound effect. The total demand is the sum of the demands of all individuals who are contemplating buying the good in question. For the sake of a simple concrete example, let us assume that in a certain village four families

only¹ are interested in the purchase of strawberries, and let us assume that we are endowed with knowledge of their respective demands as shown in the following schedule.

DEMAND FOR STRAWBERRIES, RETAIL MARKET OF X — MAY 1, 1939

<i>Price per quart</i>	<i>Quantities people are ready to buy</i>				
	FAMILY A	FAMILY B	FAMILY C	FAMILY D	TOTAL
55 cents	10 qts.	2 qts.	1 qt.	0 qts.	13 qts.
50 cents	10 qts.	3 qts.	1 qt.	0 qts.	14 qts.
45 cents	10 qts.	4 qts.	2 qts.	0 qts.	16 qts.
40 cents	10 qts.	5 qts.	2 qts.	1 qt.	18 qts.
35 cents	10 qts.	6 qts.	3 qts.	1 qt.	20 qts.
30 cents	10 qts.	6 qts.	4 qts.	2 qts.	22 qts.
25 cents	10 qts.	6 qts.	5 qts.	3 qts.	24 qts.

Family A is evidently well-to-do; they will take all the strawberries they want, regardless of what the price may be (within the limits of possible prices). To the other families the matter is more serious; their purchases will be governed by the price. The schedule in the last column is the total demand for strawberries in this simplified market.

Definition of demand. It is total demand which operates as one of the chief determinants of price. Hereafter the term demand, where not qualified, will be understood to mean the total demand.

DEMAND FOR STRAWBERRIES, RETAIL MARKET OF X — MAY 1, 1939

<i>Prices</i>	<i>Quantities</i> ²
55 cents	95 quarts
50 cents	100 quarts
45 cents	115 quarts
40 cents	150 quarts
35 cents	200 quarts
30 cents	300 quarts
25 cents	500 quarts

¹ The assumption of this small number is of course artificial. It serves to simplify the example without affecting its validity as an illustration of the general principle.

² Hereafter the word "quantities" in any demand schedule indicates quantities which people are ready to buy at various possible prices.

It may be defined as follows: *The demand for any good is a schedule of the respective quantities of that good which people are ready to buy at all possible prices.*

For the sake of a concrete example, let us expand our illustration of the strawberry market of X by including all of the hundreds of families which may be interested in the purchase of strawberries, giving a demand schedule in general character as shown in the previous table.

In Figure 4, this schedule is represented in graphical form, the curve *ag* being the corresponding demand curve.

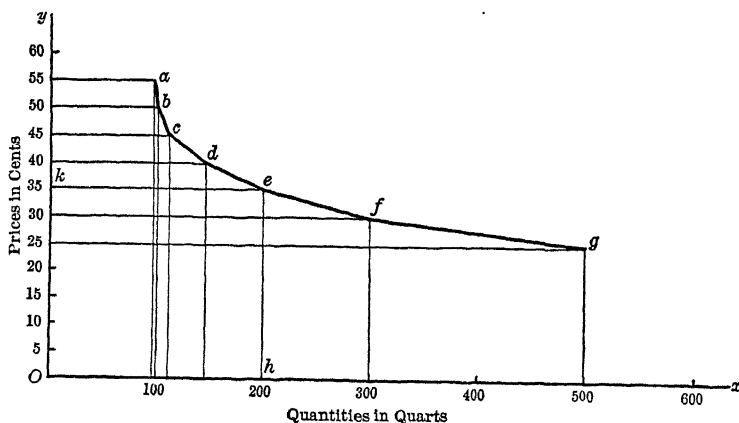


FIG. 4

Time and place of the demand schedule. The demand schedule is a series of simultaneous possibilities, only one of which can be realized at a given time. It relates to a particular date or period — to the day of May 1, 1939, or rather to those hours on that day during which buying can take place. The prices are alternative prices; only one of them can turn out to be the actual price. So of the quantities; only one can prove to be the amount actually bought. Which of the various amounts is actually bought we may not know. If we do, we must not forget that at the same time there is an ability and willingness to buy other amounts at other prices. The demand schedule is neither history nor prophecy. It does not show a series of purchases which have taken place at such and such prices. It does not show how much will be bought at any particular price

after the given date or period. Each day will have a demand schedule of its own, which may or may not be the same as this one.

In the strictest sense the demand schedule does not relate to a *period* of time at all, but merely to an *instant* of time. Theoretically conditions of demand may be changing continuously, so that each succeeding instant has its own demand schedule. For most practical purposes however it is sufficiently accurate to consider demand with reference to a considerable period, during which a certain price might hold and a certain quantity be bought. This period, be it remembered, may be quite short, often less than a day or even less than an hour. In the market for certain corporation stocks or certain commodities, such as wheat, cotton, etc., where trading is very active, conditions of demand change rapidly, and there is close approach to the theoretical condition of the constantly changing demand schedule. The demand schedule thus relates to a particular time, whether a mere instant or a considerable period, and is not a record of a historical sequence of prices and quantities bought.

Of course such a schedule must relate also to a particular place. Market conditions vary, and a schedule showing the quantities which would be bought at various prices in one market will not hold of any other market, except as a coincidence.

Limitations of this sort of analysis. The given demand schedule represents demand only in limited fashion. It does not show in what manner people's ability and willingness to buy strawberries will be displayed in case there are changes in other prices, for example, the price of raspberries. The schedule is valid only on the assumption that other prices are either not affected by variations in the price of the good in question or, if affected, bear some constant relation to the price of the good in question. In the most profound sense total demand, like individual demand, is a *schedule of schedules*, which takes into account not only all the possible variations in the price of the good in question, but also all possible variations in other prices. Demand in this sense however is a problem for more advanced study. Actually there are cases in which the price of one commodity or service is subject to variation with little or no immediate change in other prices, and there are many other cases in which other prices follow the variations of a particular price

without change in their comparative relations. Thus our simple demand schedule is not without practical significance. Like the scientist who first isolates certain chemical elements in a pure state in order to comprehend more complicated phenomena, so we concentrate attention on this simple type of demand curve as a step toward understanding the more intricate workings of demand in general.

Demand and desire. Demand is not the same as desire or need. An impecunious student, standing with empty pockets outside a theatre, may have a keen desire to see the show, but we cannot say that he has a demand for a ticket. His desire has no effect upon the demand for tickets, no effect upon the numbers that could be sold at various prices. A starving man needs food, but if he cannot buy any, he offers no demand. To have a demand for anything one must not only desire it; he must also be able and willing to buy. Most of us desire fine automobiles, trips to Europe, and other things without number. Most of these desires have no effect upon the demand for these things. The only one who contributes to the demand for a certain make of automobile is the one who is able and willing to buy at least one at a price at which this automobile is likely to be for sale. We have already paid some attention to the interesting subject of human desires or wants¹ and have seen how closely they are related to demand. Though human wants are obviously the origin of the demand for any good, we need to be on our guard against thinking desire and demand are the same thing.

The law of demand. If the reader will glance at Figure 4, he will notice that the demand curve in this chart slopes downward as it goes from left to right. As we move along the curve in this direction, the distances that measure prices grow shorter while the distances that measure quantities grow longer. In graphic form this is an illustration of an important principle known as the *law of demand*. The principle may be stated as follows: *In general, the quantity of any good which people are ready to buy varies inversely with the price of that good.*²

¹ See Chapter I.

² As we shall use the terms, "inversely" and "directly" do not imply that the variation must necessarily be proportional. If a increases when b increases or decreases when b decreases, we say that a varies directly with b , even though the

This law has been developed with the aid of numerical schedules and graphs, and these devices will continue to serve us in the further pursuit of the laws of price. The reader hardly needs to be reminded that the exact quantities used in these schedules are a matter of presumption and are not significant except as a means of illustration. The general relation between price and quantity — *i.e.*, that the quantity people are able and willing to buy increases with a decrease of price and *vice versa* — is a truth beyond question. It is only when we undertake to express this in exact numbers that we must deal in presumption. In actual business it is seldom if ever possible thus to give exact values to the demand schedule, although the business man is perfectly certain of the general relation between price and the quantity that people are ready to buy.

Special cases. We have used the phrase “in general” in the above generalization in recognition of certain special cases in which there is a different relation between price and the quantity that people would buy. If the price of diamonds, for example, should fall so low as to place them within the reach of everybody, it is quite possible that sales, instead of being larger or even as large as they are, would actually be smaller. Desire to have diamonds is stimulated to a large extent by the very fact that they are high priced. If everybody could buy them, probably few people would want them. At times when people have plenty of money, the same perversity is sometimes illustrated in the sale of other luxuries and novelties. For example wood carvings from Switzerland have sometimes been found to sell much better at a price considerably above the level which would represent a normal return on cost of importation including duty. Again it happens not infrequently that people, not knowing much about the quality of some good, assume that the higher priced article must be better than one which is lower priced and so are more inclined to buy at higher prices than at lower prices. Merchants have many stories to tell of such peculiarities of the

change in *a* may not be exactly proportional to the change in *b*; and correspondingly of an inverse variation.

In accordance with strict mathematical terminology it would be more correct to say that the quantity people are able and willing to buy is “a decreasing function” of the price; but this phrase has been found to be awkward and cumbersome in ordinary speech and calculation. The principle as stated above has abundant precedent in good literary usage.

buying public.¹ Demand under such conditions may be represented as in Figure 5. Such occurrences are more frequent in boom periods than in times of depression and of course do not represent the ordinary reaction to price.

There is another group of special cases which arises out of perfectly rational action. Suppose a state calls for bids to erect a new

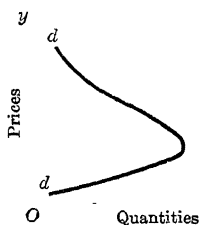


FIG. 5

capitol building. Here there is only one buyer in the market, and there is only one article to be exchanged in any case. The state will take this one capitol building, no more and no less, regardless of the price, up to the limit where the price is so high as to cause the state to give up the building project altogether. What sort of demand schedule and demand

curve do we have in this case? In the price column we will list all prices from the lowest that any contractor is likely to bid up to the highest which the state will pay. Opposite each of these figures in the column of quantities will of course appear the figure 1. The corresponding demand curve will evidently be a straight line parallel to the y axis and stopping at the highest price the state would be willing to pay.

The same sort of demand appears when a single buyer calls for bids for a fixed quantity of a certain special article for which there is no demand elsewhere. For example a national government might need 500,000 gas masks of a particular style for its army. It calls for bids. The quantity taken will be exactly 500,000, whatever the price may be, and we have again the vertical demand curve.

These special cases we must recognize when we come upon them, but they need not blind us to the existence of the general principle or tendency.

Explanation of the law of demand. Some people are content to take the general principle of demand as an observed fact without further inquiry. Our investigation of marginal utilities however makes it possible for us to understand the foundation of the prin-

¹ Cf. Dane Yorke, "The Mystery of Retail Price," *The American Mercury*, December, 1933, especially p. 391.

ciple. It is not entirely a matter of psychology, since demand depends not merely on desire but also on ability to pay; that is to say, it depends upon an individual's resources and on the way in which he wishes to distribute them among alternative uses. An increase in price not only makes a good less attractive in comparison with others, it also makes it physically impossible for people to buy as much as they could and would at lower prices; money does not go so far. A decline in price has an opposite effect purely as a matter of arithmetic. For explanation of the way in which individuals distribute their income among different uses, we must hark back to the balancing of relative marginal utilities. If the relative marginal utilities of two goods do not stand in the same ratio as their prices, an individual will have reason to change the way in which he spends his money. If the price of a good rises he will restrict his purchases of it until its relative marginal utility has increased by an equivalent amount. On the other hand, if the price falls, he will have reason to expand his purchases so that a general balance can be effected. These considerations people may sense very vaguely; they give them no technical names, but their response is unmistakable. The result is the law of demand.

We observe then that the law of demand, as exemplified in any particular commodity or service, depends upon a multiple operation of the law of marginal utility. The relative marginal utilities of this particular good to various individuals vary inversely with the respective quantities they may consider. At the same time the relative marginal utilities of other goods are subject to the same principle. As buyers increase their expenditure on one good, its relative marginal utility to each of them decreases. Spending more on one thing, they have to cut down their purchases of others; and the relative marginal utilities of these other things to each of them increase. Thus each and every buyer is subject to restraint by the counter-working of the same psychological principle. This restraint is intensified by a high price and diminished by a low price; its outward manifestation we see in the law of demand.

As a further practical consideration it is important to note possibilities of complete substitution. If the price of butter rises too high, some people are prepared to use oleomargarine entirely. On

the other hand, if lamb chops were only as cheap as Hamburg steak, what a change there would be in many a meat platter! The operation of the law of demand as regards any particular commodity or service may indeed be considered essentially in terms of the availability of substitutes. If a small quantity only of a given good can be sold at a high price, the chief reason may be that, subject to this condition, people prefer to buy substitutes. On the other hand, if a large quantity of a given good can be sold at a low price, the chief reason may be that, subject to this condition, people are able and willing to buy this particular good in place of many other possible substitutes. In a broad sense all goods are substitutes for one another. When an individual is trying to make up his mind whether to spend more money on clothes and less on a summer vacation, he is comparing two kinds of "goods" which are in no sense similar. Nevertheless they are substitutes in the sense that his desire for one limits his willingness to spend money on the other. However, although these very dissimilar things must be regarded as substitutes in a broad sense, the demand for a particular brand of suits is likely to be affected intimately only by its availability and the price of close substitutes, such as another brand.

Elasticity of demand. Every household uses a certain amount of salt. It would be most uncomfortable to go without salt or even to cut down appreciably upon the usual amount consumed, yet there would be little to be gained by increasing the amount used, since everybody already has about all he wants. If the price of salt should be unusually high, people would still buy nearly as much as usual. The amount expended for salt is a very small part of the household budget, and even at a much higher price, the cost would be small and the increase could be made up by a slight saving on some other item of expenditure. On the other hand, a very low price of salt would not mean any great increase in the quantity bought, for the simple reason that people were already buying about all they wanted. The law of demand holds good of salt, as of other things; that is, at any given time the quantity that people are able and willing to buy varies inversely with the price. But it varies only slightly. In such a case it is said that demand is very inelastic. The demand curve has a steep slope as shown in Figure 6. Changes

in the length of the lines measuring price (as Oa , Oc , etc.) are accompanied of course by opposite changes in the length of the lines meas-

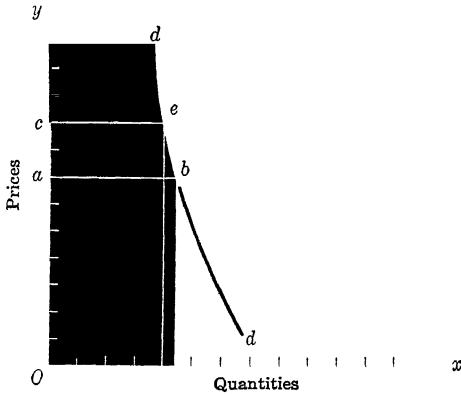


FIG. 6

uring quantities (as ab , ce , etc.), but these changes are comparatively small.

On the other hand, there are things which have an elastic demand. This means that variations in price are accompanied by comparatively great variations in the quantities that buyers would take. If the price of a certain

high-grade automobile were \$500 instead of \$3,000, the number of persons able and willing to buy that particular make of automobile would be enormously increased. If the price were \$10,000, the number of purchasers would be only a small fraction of those who would buy at \$3,000. The demand for any particular kind of

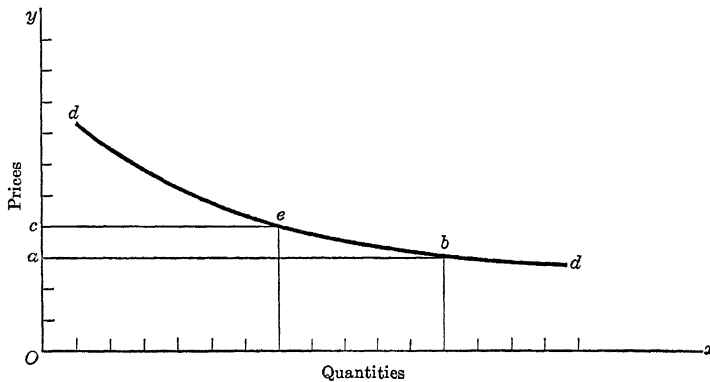


FIG. 7

car is extremely elastic. The demand curve in such a case has a gentle slope as illustrated in Figure 7. This is because changes in the distances measuring prices (as Oa , Oc , etc.) are accompanied by

relatively great changes in the distances measuring quantities (as ab , ce , etc.).

Degrees of elasticity. In order to distinguish between different degrees of elastic and inelastic demand, it has been found convenient to set up a conventional dividing line; namely, that demand schedule in which the quantity taken changes inversely in exact proportion to the corresponding changes in price. In other words, the amount of money spent, being the product of the quantity taken and the price, is always the same. Such a demand is illustrated by the curve dd in Figure 8. Whatever point be taken upon

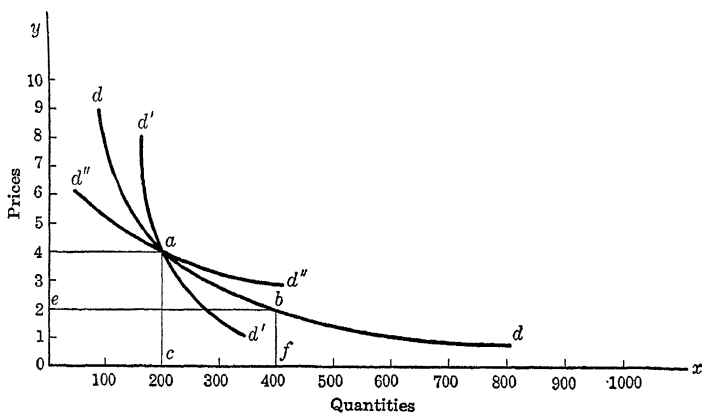


FIG. 8

this curve, as a or b , the product of the vertical height from the x axis and the horizontal distance from the y axis, measuring price and quantity respectively, will always be 800.¹ The reader will note that these products are represented on the diagram by rectangles, whose areas are equal. If the demand is represented by such a curve, it is sometimes said that demand is neither elastic nor inelastic. If the demand is such as to show a steeper curve, as $d' d'$, it may be said that demand is inelastic. A curve less steep, as $d'' d''$, would indicate an elastic demand.

It is an interesting principle that, when this criterion of elasticity is used, an inelastic demand for any good indicates a condition in

¹ The reader who is mathematically inclined will recognize this curve as a rectangular hyperbola, whose equation is $xy = a$.

which the total amount of money that people are ready to spend for that good varies directly with the price, and an elastic demand exists when the total amount of money so spent varies directly with the quantity bought.

The limits of elasticity. The absolute limit of inelastic demand is reached when the quantity people are able and willing to buy is the same irrespective of price. Such demand is represented graphically by a vertical straight line. The case of the sole buyer, such as a government calling for bids for a definite quantity of a certain special commodity, is an example. Here it is obvious that total expenditure varies directly with price. In actual practice however absolute inelasticity of demand is possible only within a given range of prices.

The absolute limit of elastic demand is reached when buyers are ready to buy any quantity at a given price but would buy nothing at any higher price. In other words, the price people are able and willing to pay is the same irrespective of quantity. As an example we may cite a large oil company which agrees to purchase all the crude petroleum brought to it at a posted price. Demand in this case is represented by a straight horizontal line parallel to the horizontal axis and at a height indicated by the determined price on the vertical axis. Here it is obvious that total expenditure varies directly with quantity. In actual practice this is possible only within a given range of quantities. While it is conceivable that this extreme condition of absolutely elastic demand might exist with two or more independent buyers, such cases are seldom found in practice. What makes demand reach this extreme is generally the presence of a single buyer who is alone ready to buy the good in question in the particular market and to buy all that is likely to be offered for sale.

Examples of inelastic demand. As a general rule inelastic demand will be found in the case of simple necessities, since people must have a certain amount, even at the cost of heavy sacrifice of other enjoyments, and since after their needs are fairly satisfied they do not much care for an increased amount. Salt is an example of a necessity. Inelastic demand is also found in the case of an article for which the desire may be fully satisfied by a small quan-

tity. Salt illustrates this characteristic also, but other illustrations may be chosen from goods which are not also necessities. Consider, for example, the article with which young women improve upon the color imparted by nature to their lips. The quantity of lip-sticks purchased depends upon fashion and personal or family taste; it is not much governed by the price. Few who are inclined to use this beautifier would be deterred by a price even considerably above the prevailing one. And few would be led to its more copious use even though the price should become extraordinarily low. Of course the reason is that anyone can completely satisfy her need with an expenditure of only a very small part of her income, no matter what the price. The demand is thus inelastic.

A more practical illustration of inelastic demand is exhibited in the following record :

PRICES AND CONSUMPTION OF POTATOES IN THE UNITED STATES ¹

<i>Year</i>	<i>Average wholesale price per bushel for the crop year</i>	<i>Per capita consumption in bushels</i>
1907	\$0.61	3.68
1908	.79	3.47
1909	.38	4.34
1910	.48	3.76
1911	1.03	3.26
1912	.49	4.40
1913	.65	3.45

The reader will not have failed to observe that the above table is a record of historical sequence and not a picture of amounts that would be consumed under different assumptions as to price at a given time, as is required by the definitions of demand and elasticity of demand. It is quite safe to infer however that the comparatively steady consumption of potatoes in face of a wide fluctuation in price is due to the fact that potatoes are an article of relatively inelastic demand.

Examples of elastic demand. Substitution. The demand is generally elastic for luxuries, things that people could do without, especially if the article is also expensive. The theatre is a luxury. The number of tickets that can be sold varies greatly with the price

¹ F. C. Mills, *The Behavior of Prices*, 1927, p. 148.

of admission. Witness the immense increase of attendance upon the theatres since the "movies" have made it possible to offer entertainments at prices far below what was possible for the old style show. The demand for washing machines is also elastic. The machine is expensive. Many a family cannot afford one. But the utility of the washing machine is so great and so well recognized that a material reduction in price would certainly lead to a great increase in the numbers purchased. On the other hand, an increase in price would soon put the machines beyond the means of many who now purchase them.

We have already noted how the operation of the law of demand is accentuated by the availability of substitutes. The degree of accentuation is largely measured by elasticity of demand. The demand for a general line of product (say gasoline) may be fairly inelastic, but the demand for any particular brand of this product is generally very elastic. People may have some preference or other for particular brands, but slight differences in price will swing many of them away from ordinary choices. Suppose the price of Texaco gasoline should be kept one cent per gallon higher than other brands. How many followers of the Fire Chief would keep on buying it? In like manner the demand for a particular brand of almost any product is extremely elastic. The demand for salt in general is very inelastic, as we have observed; but the demand for particular brands of salt is extremely elastic. This is not only true of particular brands, but also of particular kinds of product. The demand for veal is elastic because people also have beef, pork, mutton, and lamb to choose from. For meat in general there are many substitutes; for example, fish or eggs. These substitutes increase the elasticity of demand for meat.

Necessities and luxuries. The terms "necessity" and "luxury" are dangerous ones to use in this connection, without careful definition. If by necessity we mean a thing which is required to sustain life or to maintain a decent living in accordance with prevailing standards, then the demand for necessities will generally be inelastic. But it must be recognized that the demand is just as inelastic for certain things which are not at all necessities under this definition, and this does not refer merely to such trivial luxuries as the

lip-stick. Take for example tobacco, which certainly is not a necessity, as we have just defined the term. Yet few articles of common use show so inelastic a demand. The ordinary man will insist upon having his smoke, no matter what the price. If necessary he will make almost any sacrifice in the other items of his personal or family expenditure. The same is often true of liquor. If a luxury is to be regarded as the antithesis of necessity as we have defined the latter term, then the demand for luxuries will as a general rule be elastic. But there will be many and important exceptions, of which some examples have already been noted. Indeed the distinction between "necessity" and "luxury" is at best a vague one, and this distinction is far less useful for the scientific analysis of demand and price than the distinctions which rest upon the degree of elasticity of demand. On the latter basis we may throw tobacco and lip-stick into the same category as bread and meat, without raising the question of whether anything is a necessity.

EXERCISES

1. Draw up an assumed schedule of the demand for potatoes on a certain day in a small town.
 - (a) Construct the demand curve corresponding to this demand schedule.
 - (b) Take any point on the demand curve, and draw perpendiculars from this point to the horizontal and vertical axes. Exactly what does each perpendicular represent?
2. Describe the essential features of a demand schedule. What factors may cause a change in demand from one time to another? Draw a simple diagram to illustrate an increase in demand; a decrease in demand.
3. Draw up a demand schedule for each of two commodities, one of which has an elastic demand, the other an inelastic demand. Test each demand schedule to insure that it conforms to the requirements. Plot the two demand curves, on separate graphs.
4. Explain how each of the following factors tends to affect elasticity of demand:
 - (a) Necessity or luxury
 - (b) Presence or absence of substitutes
 - (c) Habitual consumption
 - (d) Size of the expenditure for the particular good as compared with one's total expenditure
 - (e) Advertising warning the buyer to "beware of imitations" or "accept no substitute."

-
5. In each of the following cases, what condition is implied with respect to elasticity of demand? Explain.
- (a) On several occasions in recent years the American Telephone and Telegraph Company has reduced its toll rates for evening calls.
 - (b) A small wheat crop may result in a greater income to wheat growers than a large crop.
 - (c) In 1936 passenger rates on American railroads were reduced from 3 cents to 2 cents per mile, at a time when the railroads were in dire need of additional revenue.
 - (d) The Brazilian government has frequently resorted to burning coffee or dumping it in the ocean, in an effort to increase the incomes of coffee producers.
6. (a) A city calls for bids to furnish it 50 garbage trucks of a particular type. Draw the appropriate demand curve.
- (b) The United States Treasury stands ready to purchase all gold brought to it at \$35 per ounce. Draw the appropriate demand curve.
7. Does the schedule for Family A on page 161 contradict the principle of diminishing marginal utility? Explain.
8. (a) Construct a schedule and a diagram demonstrating the principle, stated in the text, that when demand is inelastic the total amount of money that people are ready to spend varies directly with the price.
- (b) Present corresponding demonstration for the condition of elastic demand.

IX

DEMAND AS IT CONCERNS THE SELLER

We have thus far been studying demand from the viewpoint of the consumers, the purchasers of goods. Demand emanates from the consumers; it is the force they bring to bear on the market. As such, demand becomes a matter of vital concern to all those who are engaged in the selling of goods. In order to understand the nature of demand as it concerns the sellers, we need first to know what a market is and to appreciate the characteristics of different kinds of markets.

The market: Definition. The word "market," in the vocabulary of economics, has a somewhat broader content than is given it in ordinary historical descriptions, such as have appeared in Part I of this book. As technically defined, *a market for any good is a place where buyers and sellers exchange that good.* This includes obviously all those markets which we have been studying. It includes also every place where buyers and sellers exchange goods even though there may be no formal market organization. Thus the retail strawberry market of a city need not be any formally organized market in a special building or place; it may be simply the various retail stores and offices where dealers and consumers sell and buy strawberries. It is not even necessary that all the buyers and sellers be physically present at the market. They may be in communication with the market by telephone, telegraph, or mail and represented there by agents or brokers. The market is simply the place at which buyers and sellers make their influence felt, at which exchanges are made, the meeting point of the forces that determine price.

Free Market. *A free market is a market in which there is no rationing of goods and no compulsion on either buyers or sellers to go through with transactions which they do not wish to carry out.* In our following discussion it will be assumed, unless otherwise stated, that the term market means a free market as thus defined.

Who sets the price? Although our definition of the market is broad enough to include practically anything that would be thought of as a market in the common usage of the term, it will be helpful for our purposes to distinguish several different kinds of markets. Economists customarily speak of "the prices established by the market," but it may well occur to the reader that a market is a human institution, made up in the last analysis of people, and that when we speak of a price being determined by the market, we mean that the individual or individuals who set the price have been induced to set it at some particular level by forces which operate on them in the market. In any economic study the fact that prices are always set by someone should never be overlooked. We shall find it convenient therefore to base our first classification of markets upon the different ways in which prices are actually set.

In most markets in modern industrial nations price is set by the sellers. If you, as a consumer, enter either a large department store or a small grocery store, you will find in every case that there is a certain established price for each article in the store. You are free to buy as much or as little at that price as you choose, but it is rare indeed that you will be able by your own unaided action to secure a change in the price.

This kind of market we may contrast with those markets where the prices are named by the buyers. When the large milling companies and grain dealers purchase wheat from the farmers throughout the western United States, it is the buyers who set prices. The farmer coming in with a truckload of wheat to the elevator can either accept the price which he is offered or take his wheat home. He has the option of selling as much or as little at that price as he wishes.

In both of these cases it is of course conceivable that the individual buyer in the store or the individual farmer might be able to secure some change in price and would thus have some effect upon prices through his own actions, but even so it is always the seller, in the first case, and the buyer, in the second case, who actually names the price at which transactions are made.

There is a third type of market which differs from both of these patterns. In the wheat pit or in the stock market or even in dealings

between one buyer and one seller, price is apt to be decided by a process of bidding and counterbidding. A buyer may go into the wheat pit and offer to purchase a thousand bushels at 68 cents a bushel. A seller may offer to sell a thousand bushels at 72 cents. Either party may name a price and in such a highly organized market there will always be sellers willing to make sales at a price just above the current market price and buyers willing to purchase at a price just below it. The bidding and counterbidding will of course take a different form if, for example, the president of a manufacturing company is negotiating for the purchase of a piece of special machinery from a manufacturer of machinery. But here again each party will name a price at which he would be willing to make the transaction, and the price finally agreed upon will be the result of their bargaining back and forth. We may thus distinguish three sorts of markets: those in which the buyers name prices, those in which sellers name prices, and those in which either party may name a price.

Naming either quantity or price. This mention of the different kinds of markets suggests one characteristic which they all possess in common. It is true of any free market that an individual, whether he be buyer or seller, may name either the quantity or the price which will govern a transaction but cannot name both. For instance it is scarcely conceivable that a consumer should go into a department store and be told that he must pay a dollar apiece for golf balls and that he must buy a dozen of them, or that the farmer bringing his wheat to the elevator should be told that he will receive 72 cents a bushel and must deliver five hundred bushels.

Of course the store may very well refuse to sell golf balls except by the dozen or the elevator to buy wheat except in lots of one hundred bushels, but in these cases the firm which names the price is merely establishing the unit in which it chooses to deal. The principle stated above is not disturbed by the fact that the individual who names a price may also name the unit of measure.

This principle operates in the third type of market as well as in the first two. If an investor wishes, he may order his broker to sell one hundred shares of stock at the market price or he may order his broker to offer one hundred shares for sale at, say, \$58.00 per share.

What he does not have the power to do is to insist that one hundred shares shall be sold and that they shall be sold for \$58.00 per share.

In contrast we may note a familiar situation where this principle does not operate. If you go into a department store the day before Christmas, you may very well be told that they "are out" of something which you very much wish to buy. In this case you, as buyer, have the right to name neither price nor quantity. But this is not an exception to the principle, it merely means that there is not here a free market for that particular article. When the store is unable or unwilling to sell something which you desire and which it normally carries, the supply of that article is actually being rationed, just as sugar was rationed during the World War.

Competition: Definition. From the seller's viewpoint, the most significant classification of markets relates to the degree of competition which prevails between himself and other sellers. *Competition is the effort of two or more sellers to sell to the same person or persons, each seller acting independently in his own interest without regard to the interests of other sellers.*¹

Competition and monopoly. Degrees of competition. The antithesis of competition is monopoly, a topic to which we shall give attention in a later chapter. It will be at once evident that the absolute extremes of either perfect competition or perfect monopoly are seldom found in the world of today.² The actual condition, under any given practical circumstances, is almost without exception somewhere between these extremes. In dealing with practical problems it is therefore seldom possible to use the terms competition and monopoly in any rigid sense. These are relative terms, indicating a degree of approach to one or the other extreme. Various qualifying terms have thus arisen in the literature of

¹ This is obviously competition of sellers, and there may be also competition of buyers, with an analogous definition. The latter concept is however seldom employed, and where the term competition is used without qualifying words, competition of sellers is to be understood.

² In strict theory there is not perfect competition unless a number of rigid requirements are fully met: (1) The total number of competitors must be indefinitely large. (2) Each competitor must sell exactly the same product, and it must be so regarded by all buyers. (3) Each competitor's stock must be so small in comparison with the whole group of competitors that no change in his output has any effect on the price. (4) Competition must be so keen that if any individual seller raises his price above the others, he will lose all his business.

economics. "Pure competition" is used to describe a state that approaches the competitive extreme but is less absolute than perfect competition or is "perfect" in only one respect while falling short of perfection in other respects. "Imperfect competition," if used strictly as the antithesis of perfect competition, would of course be synonymous with monopoly. As the term is generally used however, it denotes a state which may still be regarded as competition, though it is far "less perfect" than pure competition. The term "monopolistic competition" is sometimes used to refer to a particular kind of imperfect competition. The rather forbidding term "oligopoly" is by some writers employed to describe a condition of competition that is imperfect in another respect; *i.e.*, in that there are only a few competitors.

For the purposes of this elementary book two stages only, between the extremes of perfect competition and perfect monopoly, will be sufficient; for these we chose to use the terms *pure competition* and *imperfect competition*.

Pure competition: Market conditions. Definition. Pure competition is characteristic of those markets in which the buyer normally sets the price and of the great speculative markets where prices are determined by the bidding of buyers and sellers. The essential features of the first type of market are as follows. The article dealt in is a standardized product. All units offered by the different competitors are exactly alike, and no buyer has any reason for preferring to deal with any particular seller. The number of sellers is very large, and the quantity that could be offered for sale by any seller is a very small part of the total quantity offered.

This situation is illustrated by the regular markets for many great commodities, such as wheat, cotton, and certain other staple agricultural products. We do not usually think of individual wheat farmers or cotton growers as competing with one another. Yet inasmuch as they are all selling a standard commodity, they are competitors as truly as two manufacturers each of whom sells electrical refrigerators or automobiles. What obscures the situation is the fact that there are so many of them that each one has a negligible influence on the prices established on the market and so on the fortunes of his fellows. The largest wheat grower in the world does

not produce more than a small fraction of the world output of wheat.

This situation generally exists also in the great speculative markets where dealings take place in standard commodities, such as wheat and cotton, and in corporation securities and government bonds. Here, it is true, a large producer or dealer can sometimes temporarily influence the price on account of his large offers. But except for occasional short periods, the price of wheat at the Chicago Board of Trade, as well as the price of General Electric Company stock or United States treasury notes of 1945 on the New York Stock Exchange, is beyond the control of individual sellers.

Pure competition may be defined as *competition between sellers under such conditions that the individual seller is unable by his own actions appreciably to influence the market price.*

The individual seller meets perfect elasticity of demand. The position of the individual seller under conditions of pure competition may be described in terms of the demand for his product. The

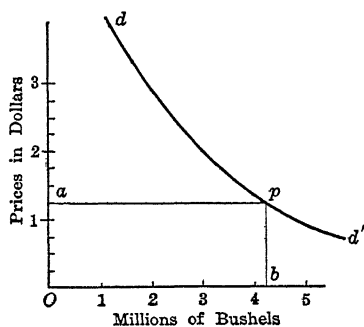


FIG. 9

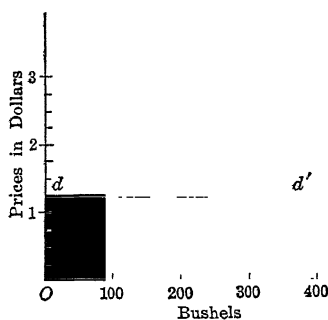


FIG. 10

farmer comes to market and finds that the price of wheat is \$1.25 a bushel. At that price (or by shading it ever so slightly) he can sell any amount of wheat he wishes, and he will not be able to sell any wheat at all at a price even a small fraction of a cent higher. His sales will not affect the price. Whereas the total demand for wheat in the market at this time may be represented by the normal downward sloping curve, as in Figure 9, the demand for the wheat of the particular farmer is represented by a horizontal line, as in

Figure 10. It will be recalled that such a line is the graphical representation of perfect elasticity of demand. And it is strictly true that the demand for the product of the individual seller in such a market is completely elastic. This is simply another way of saying that the seller has to accept the market price as he finds it.

This will be true for any producer who sells his product in a market of which he is so small a part that he cannot influence the ruling price. It will often be approximately true even when this condition is not fully met. In the industrial field the elasticity of demand for the output of any one producer will be affected by the similarity of the products of other firms to his own. Where the products of different producers have a high degree of individuality, as may be illustrated by the three closest competitors in the automobile market, the Ford, the Chevrolet, and the Plymouth, the demand for any one of them will certainly not be completely elastic. But there are industries in which the products of different sellers, while not identical, are still hardly distinguishable. Again there are certain industries where, although the number of producers may be small enough and certain ones large enough to give an individual seller occasionally some influence on the price, it is still correct to say that each seller is generally confronted with a perfectly elastic demand for his own product.

It should be noted that the elasticity of demand for the individual producer's output has very little to do with the elasticity of total demand for the commodity which he produces. It is obvious that, although the demand for food stuffs and especially for wheat is highly inelastic, yet there is a perfectly elastic demand for the output from any one farm.

The seller has no "price problem." Whereas the manufacturer of almost any branded or trade-marked product, such as a washing machine or a particular kind of corn flakes, must before setting his price think carefully of the probable consequences on the amount of his sales, the wheat farmer has no such "price problem." He does not have the problem of deciding to set a price of, say, \$1.27 on the wheat he brings into the grain elevator. Of course, since he sells in a free market, he could be said to be free to set the price at which he would sell instead of the quantity. But this freedom would be

rather illusory, because all it means is that he could set his price slightly above the market price if he were willing to sell none of his wheat or below the market price if he were willing to forego voluntarily a part of the money which he might receive. The reason, in other words, that he does not choose to "set the price" is that there is no conceivable advantage to him in "setting" a price which differs from the established market price.

Imperfect competition: Inertia and ignorance. Once the characteristics of pure competition are understood, it will be easily recognized that in most markets competition is generally not pure. There are many factors which in actual practice prevent the attainment of pure competition.

In the first place, if a market is not well organized, if buyers and sellers have difficulty in coming together and comparing each other's wares and each other's prices, we have imperfect competition. Competition in the retail business is often imperfect to a pronounced degree, because it is so much trouble for buyers to get around and find out exactly what each store is charging for each and every article. Furthermore it is often difficult for buyers to determine how the quality of goods offered by one store compares with the quality in other stores. Competition in the wholesale business and in the great speculative markets much more nearly conforms with the analytical idea of pure competition, chiefly because the buyers are better informed and more energetic to seize any advantage.

Individuality of products. Competition is imperfect in the technical sense also when the products of competing sellers are not precisely alike. It is typical of the field of industry, as opposed to that of agriculture, that the products of particular producers have considerable individuality. The individual producer is not selling grade A wheat or even a standard quality of cement; on the contrary he is selling a particular kind of machinery or a particular brand of beer or a particular make of automobile. Wherever this is true, individual consumers will have preferences for one make or another. Each producer will have a number of regular and loyal customers who can be depended upon to buy his product even if its price is somewhat higher than others of similar quality. If the price

of one make falls a very little below that of another, all buyers will not desert the other makes in a body and concentrate upon the one with the slightly lower price. In the automobile industry, for example, there are many peculiarities in detail among the three close competitors, Ford, Chevrolet, and Plymouth. It is obvious that the demand for Mr. Ford's car will depend very directly upon the price that is being charged for Chevrolet and Plymouth. If the prices of Chevrolet and Plymouth are cut by \$20.00 a car and if Mr. Ford does not follow suit, they will take a certain amount of his business away from him. But it is unthinkable that such a price cut would completely deprive the Ford Company of its customers. In this respect the position of the automobile manufacturer offers a sharp contrast to that of the wheat farmer.

Trade marks and special service. Even though the products of different manufacturers may be virtually identical, the sellers frequently succeed in creating the impression of differentiation in the minds of the buyers, especially in the retail markets. Each sugar manufacturer puts up his product in a different package and claims for it distinctive advantages. Canned goods have various labels and various contents. Meats, clothes, gasoline, and very many other everyday products in modern America are offered by each competing producer in a form which he tries to individualize so that it will be known as his own. Brands, forms of packaging, distinctive coloring, slogans, all sorts of advertising are some of the devices used to gain this effect and so to attract a particular following among the buying public.

Consciously or otherwise each individual store develops for better or worse an individuality in its offerings, even though it sells the same brands as another store across the street or another store around the corner or one in the next block. What people pay for in a store is not just merchandise but also the service that goes with it, the general atmosphere of the place, and the convenience of its particular location. These are some of the things which tend to make competition imperfect. As a result business becomes a conflict, not merely between people selling the same thing, but between various big and little firms each with a certain degree of monopoly power.

Small number of sellers. Even in the case of standardized industrial products like cement, it is plainly true that the individual producer cannot sell as much as he wishes at a good market price. For in these industries the number of competitors is generally small enough relative to the size of the market so that no single producer can expand his sales substantially without "spoiling the market" for his competitors and probably for himself. In such industries as these it is typically the case that a single producer will sell up to ten or fifteen per cent of the total sales in a given market. (Of course this is not the same thing as ten or fifteen per cent of the total sales in the country.) If he attempts, let us say, to double his output, it is clear that the market is going to be disturbed; he will not be able to increase his sales so largely without either engaging in an intensive campaign of advertising and sales promotion or — what is more likely — cutting his prices. Either one of these methods will provoke his competitors to retaliation. If he cuts prices they will be forced to follow, and if he organizes an aggressive sales campaign they are likely to do the same. It is of the very nature of such an industry that no one producer can expand his output very greatly without being willing to sell at lower prices or to incur much heavier selling costs or both. The individual producer is too large a part of the particular market in which he sells to have it unaffected by the quantity of which he disposes.

Demand not perfectly elastic. We observe that competition may be imperfect (1) because of the ignorance or inertia of the customers of the retail markets, or (2) because of individuality of competing products — whether actual or imaginary, or (3) where there are too few competitors to make the influence of any one of them negligible, even though they are all selling a standard commodity. When any one or more of these conditions is present in a market, the elasticity of demand for the output of a single seller will not be infinite or perfect. Henry Ford knows very well that he cannot expand sales without lowering price or increasing his selling outlays. Even the small independent grocer in a large city where the number of sellers is very large and where the products which they are selling are practically standardized is in different position from the wheat farmer. In all such cases the seller will have a "price problem," and

it is notable that virtually all of the markets in which the elasticity of demand for the individual seller's output is not absolute are ones in which the sellers name prices.

Definition of imperfect competition. The competition between sellers under these conditions is *imperfect competition*, which we define as *competition between sellers under such conditions that the individual seller is able by his own actions to exert some influence upon the market price.*

These matters adapt themselves readily to graphical illustration, as in Figure 11. Let us suppose we have an automobile manufacturer

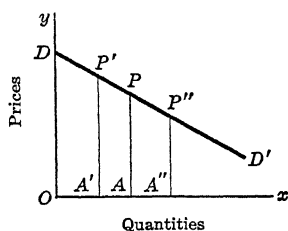


FIG. 11

turer who comes directly into competition with only three or four other makes of car. The average price of these other cars is indicated in Figure 11 by the height PA . The number of automobiles he puts on the market is determined, not by price alone, but also by his interpretation of the conditions of demand. He does not have to charge

the same price as rival manufacturers. He can charge a higher price such as $P'A'$, or a lower price, such as $P''A''$. At either price buyers can be found, as shown by the demand curve DD' . On the other hand, the presence of these other cars on the market puts a distinct limit on the height to which this particular manufacturer can successfully raise his price. Because of their presence the demand for his particular car is extremely elastic. If he raises the price to OD in Figure 11, he will have no sales. Nevertheless the fact remains that up to this level he can charge any price he wants to in accordance with his own judgment of demand, competitive conditions, and costs of production.

The liberty of imperfect competition. Although the seller's problems are more complicated when he sells under conditions of imperfect competition, another characteristic of such a situation is the comparative degree of liberty enjoyed by individual competitors. This statement may at first appear somewhat paradoxical, because it is the competition between sellers under these very conditions that is usually thought of as being the most violent. Farmers never

advertise or send salesmen into one another's territory. The economic misfortunes which befall them are usually thought of as attributable in some impersonal way to the market. But when two or three manufacturers sell their products in a market where there are no other sellers, the competition among them takes a very different form. If producer A increases his sales at the expense of producer B, it is clear to everyone what is the source of producer B's difficulties. Because the number of producers is usually much smaller than under pure competition, the competition among them takes on a more personal form and is apt to be more bitterly resented.

Nevertheless in such a market the individual seller has a freedom which he does not possess under pure competition. For him the right to set a price on what he sells is much more valuable than to the wheat farmer, for the former may raise his price, perhaps substantially, without losing all of his sales. On the other hand, the seller in imperfect competition cannot expand his sales indefinitely without lowering his price or else spending more money on advertising, salesmen's commissions, etc. As compared with the seller in pure competition, he is in a position where he has greater freedom to raise his price and where the penalties of raising it are less. On the other hand, it is much harder for him to expand his sales, and the cost of so doing is certain to be greater.

Limitations upon the seller's liberty. Although each individual competitor may have a degree of liberty in the determination of his own price, this liberty may be so restricted as to be almost negligible, or not worth using. Apparently a business man may have few rivals in his particular field; actually and indirectly, he may have many. Thus in a particular town there may be only very few stores selling a given kind of product, and competition may appear to be solely a rivalry among themselves. All of these stores however may be competing, not only with each other, but also with mail order houses and with larger stores in a nearby city. True it is that not every shopper has time or inclination to price the article he wants in every shop before he buys it; but some people do go to much trouble in this direction. It is true that every alert business man tries to individualize his product and his par-

ticular type of service. He may thus succeed in establishing more or less permanent business relations with some of his customers. But at the same time a good deal of his trade may be with casual buyers who care no more for him than for any of his rivals. To them one particular brand of merchandise is about the same as several others, and several possible places of purchase may be equally convenient. In spite of the fact that he could raise prices and still do business, the man with established customers often finds it advisable to keep prices down to a freely competitive level in order not to lose the business of alert and independent people who can be easily attracted elsewhere by lower prices.

Liberty and discretion in big business. In the matter of price it may appear that the larger business enterprises have a greater degree of liberty than is enjoyed by small businesses. Nevertheless there is need for discretion even in the management of an outright monopoly. There is the restraining influence of possible and actual substitutes, a risk which each large concern shares with its rivals; there is the danger of additional competition, the danger of stifling future demand by too high a price in the immediate present, and the possibility of all sorts of popular and political outcries. In addition the large concern which is not a monopoly is held in check by its competitors. It may still have left a considerable degree of liberty in its price policy, but this liberty has to be used with caution.

It is noteworthy that in those industries where competitors apparently have the greatest individual freedom, prices are sometimes subject to the most conservative adjustments. A big concern, though bound to no particular price by external conditions, is often forced by the very complexity of its own internal organization to adhere to a relatively stable price policy. Price changes involve various notifications to the public, the instruction of numerous employees, and all the red tape characteristic of a big organization. Price changes also complicate bookkeeping. It is no wonder then that the management of any big concern, engaged in multitudes of little transactions with the general public, prefers to make as few price changes as possible. Imagine the confusion which would result if Ford or General Motors changed Detroit f.o.b. prices every day or if retail prices in a big department store were subject to hourly

fluctuations. The relative stability of prices in industries where competition is concentrated in few hands is *ipso facto* a demonstration, not of secret agreements, nor of arbitrary monopoly power, but rather an inevitable result of the complexity of such enterprises and the need for conservative managerial discretion.

Quality competition in place of price competition. The unsettling results of too many changes in prices have a strong tendency to put competition between large concerns on the basis of quality, salesmanship, and advertising, rather than price. Thus instead of selling the same car for less money, automobile manufacturers offer, or pretend to offer, more car for the same money. Business men generally prefer this sort of competition. It calls for alertness and ceaseless activity and the expenditure of large sums of money in advertising and technical research; but the efforts which each competitor thus puts forth in his own behalf do not provoke the same quick and mutually injurious retaliation as would outright price cutting. For example, if a tire manufacturer has some degree of success in a selling campaign in which he stresses a particular process of treating rubber, he may be sure that his competitors will come forth with something of the same sort; but there the thing will stop until some other new idea crops up in the industry. On the other hand, if a price war is begun, no one can tell where it will stop or how injurious it may be to all competitors concerned.

Frequently the effect and purpose of advertising is to emphasize the individuality of different products; that is, to make the demand for an individual seller's product less elastic and competition less perfect. This is another reason for the business man's preference for non-price competition.

Efforts to control demand: Under pure and imperfect competition. The modern business man is aware that demand may be subject to outside influence, and when such is the case with respect to his own product he does not hesitate to do his utmost to modify demand in his own interest. To this end the business man employs a sales force, large or small as the case may be, to whom he pays salaries and commissions; retailers are given special "margins" as inducement to "push" a particular maker's product; exclusive agencies are set up; and finally there is the whole field of adver-

tising. Whether this is worth attempting or not depends, in the first instance, upon the degree of competition in the particular market. When a product is sold under conditions of pure competition, it will never pay individual producers to employ advertising and other devices to influence consumer taste. It would quite obviously be ridiculous for a Montana wheat farmer to advertise in a national magazine and urge consumers "to use more wheat and eat more bread." The absurdity arises from the fact that he as an individual is supplying so tiny a fraction of the supply of wheat consumed in the United States that, even if his advertising increased the sale of wheat 100,000 bushels per annum, the effect on his individual sales would be negligible. But where a producer is one of a small number of sellers in the market or is the seller of an individualized product, he may find it worthwhile to advertise.

This is one of the salient features of imperfect competition. We have had occasion before to note the impersonal character of most markets characterized by pure competition; this absence of advertising under conditions of pure competition is one aspect of that same impersonality. We have noted also that the seller who faces imperfect competition, unlike the seller in a perfectly competitive market, has a "price problem." We now see that his situation is further complicated by a "sales problem." In other words, he will have to decide not only what is the most profitable price to set, in view of the fact that a lower price will expand his sales and a higher price will check them, but also how much money he is going to spend in pushing the sales of his product.

Of course many products which at one stage in the process of production are sold under conditions of perfect competition are later sold, perhaps in a different form, in imperfectly competitive markets and are then advertised. For example, though the wheat market is one of pure competition, bread is usually sold by one or two large bakeries in each town. The bread which is baked by each firm will have its distinctive trade mark and thus will be an individualized product. Consequently bread is very generally advertised, although wheat is not. A similar example may be found in the sale of securities. A member of the New York Stock Market who receives an order from a customer to sell 10,000 shares of U. S. Steel

Common Stock would never think of advertising or even of sending salesmen around the New York financial district. The stock could not be sold in any case except at the market price, and since the stock market is a perfectly competitive one, the 10,000 shares can easily be disposed of in the open market at that price. But if the same broker had himself purchased 10,000 shares of the stock and wishes to sell them to individual customers, he would send salesmen around, and he might well advertise or incur other sorts of selling expenses.

The statement that in a market characterized by pure competition advertising does not pay refers of course to the individual seller. It does not follow that, even in case of pure competition, it may not be profitable for many sellers to engage in a joint advertising campaign. Thus the producers of wheat might conceivably, through a national organization, seek by advertising to increase the demand for wheat and so its price, to the benefit of each producer. There is a great deal of national advertising conducted by associations of producers which it would not pay any one producer to undertake. In some of these cases the product is standardized enough and the number of sellers sufficiently great so that there is approach at least to a condition of perfect competition.

Advertising. The purpose of an advertising campaign is to stimulate or even create demand for a particular product. Every resource is brought to bear to attract public attention, to arouse desire, to create new wants, to entice buyers away from rival sellers. The advertiser is aware of the law of marginal utility and of the competitive nature of human wants. He seeks therefore to increase the relative marginal utility of his product by proclaiming its many advantages, and he asserts the superiority of his product over other goods which may lay claim to the consumer's attention.

The endeavor to influence demand through advertising has become an enterprise of stupendous proportions. The following is a partial estimate of the amount of money expended on three of the

Newspapers (52 cities only)	\$1,025,421,000
Magazines	128,847,000.
Radio	<u>64,974,000</u>
Total	\$1,219,242,000

principal kinds of advertising in the United States in the first eleven months of the year 1938.¹

This is obviously far from a complete account. It does not include such important types as direct advertising, street car advertising, billboard advertising, business papers, etc.; only 52 cities are included in the newspaper advertising, and finally the record does not cover the complete year.

Although few businesses are today able to function without resort to advertising, the volume of advertising naturally varies greatly as between different lines of industry. "The advertising of automobiles; toile tarticles and medicinal preparations; food, groceries, and beverages; and cigars, cigarettes, and tobacco constituted 60.5 per cent of the total national advertising in newspapers in these cities (49 cities of the United States) in 1927."²

The advertising in the first ten months of 1938, to which reference is made above, was distributed among certain important objects as follows :

DISTRIBUTION OF CERTAIN ADVERTISING IN THE UNITED STATES
FIRST 10 MONTHS OF 1938³
(In per cent)

	<i>Radio</i>	<i>Magazines</i>
Automobiles and accessories	10.6	12.4
Clothing	0.5	5.8
Electric household equipment	0.8	2.8
Financial	0.5	3.3
Foods, food beverages, confections	33.2	16.0
House furnishings, etc.	0.1	4.7
Soap, cleansers, etc.	10.8	3.2
Office furnishings, supplies	0.0	1.4
Smoking materials	12.1	6.7
Toilet goods, medical supplies	26.6	17.3
All other	4.8	26.4
	<i>Newspaper advertising</i>	
Classified		21.6
Display, total		78.4
Automotive		3.7
Financial		1.6
General		16.0
Retail		57.1

¹ United States Bureau of Foreign and Domestic Commerce. *Survey of Current Business*, January, 1939, p. 23.

² M. T. Copeland, in *Recent Economic Changes in the United States*, 1929, p. 415.

³ *Survey of Current Business*, December, 1938, p. 23.

An index of advertising activity has recently come into favor as a barometer of business prospects. Separate indexes are published for magazines, farm papers, newspapers, radio, and outdoor advertising. Combined averages are shown below :¹

Av. 1928-32 = 100

<i>Year</i>	<i>Index</i>
1922	97.2
1923	106.9
1924	107.0
1925	112.0
1926	118.0
1927	115.8
1928	114.7
1929	120.7
1930	104.9
1931	91.9
1932	71.8
1933	65.0
1934	74.6
1935	79.3
1936	89.1
1937	94.0

1938

January	79.8
February	81.2
March	83.7
April	82.4
May	80.0
June	79.5
July	77.4
August	80.3
September	82.1
October	78.4

The business man's interest in demand. Surely no further demonstration is needed of the vital interest of the ordinary business man in the demand of consumers for the product he makes or sells. Owing to the general prevalence of imperfect competition, the salesman occupies the key position in modern business, and a knowledge of the nature of the demand for his goods is a prime qualification of salesmanship. The salesman is aware that, other things being equal, the higher his prices, the greater the income and the profits

¹ The figures for 1919 to 1934 are quoted from *Survey of Current Business*, July, 1935, p. 19. The figures for 1935 to 1937 were supplied by the Bureau of Foreign and Domestic Commerce. The figures for 1938 are from the same source, December, 1938, p. 23.

of his business; but his knowledge of the law of demand tells him that other things are not equal, that a high price may defeat its end by reducing sales, while the true road to profit may be large sales stimulated by low prices. The business man would like to know the precise form of the demand curve of his product. What will be the response of the buying public to his price policy? Are people already using about as much as they want, or is there possibility of bringing in a crowd of new users by the announcement of low prices? Is any large part of his present custom likely to drop his product if the price is raised, or may he count upon a condition of inelastic demand?

It is of even greater importance to know the absolute magnitude of the demand for his product; within any given range of prices are people ready to take large or small quantities? The fate of his business may very well turn upon his knowledge or ignorance of the amount of his goods which buyers would take at various prices. And it should be noted that the enterpriser wants knowledge of the state of demand, not only as it is at the present moment, but as it will be in the future, both immediate and more remote.

Ascertaining the facts of demand. The task of discovering the actual facts of demand is anything but easy; it is in truth perhaps the most difficult part of the technique of modern business. In spite of the difficulty however the importance of the knowledge is such that vigorous efforts are made to obtain facts throwing light on the conditions of demand. Demand is a resultant of two factors, desire and the purchasing power to make desire effective. The attempts of business men to measure and foretell demand revolve about the analysis and interpretation of business statistics which bear upon these two factors.

Indicators of consumers' desires. Desires are psychological phenomena, which do not lend themselves to quantitative measurement. It is possible however by the study of the quantities of goods which consumers have purchased over a past period to measure roughly and in a broad way changes in the consumption of various products which have been brought about by modifications in consumers' habits. For example we find that the per capita consumption of corn meal in the United States declined from .597

barrel (of 196 pounds) in 1889 to .115 barrel in 1925; the consumption of wheat flour per capita decreased in the same period from 1.142 barrels to .902 barrel. The per capita consumption of beef declined from 67.8 pounds in 1900 to 58.0 pounds in 1927, while the consumption of pork was increasing from 64.7 pounds to 68.5 pounds per capita.¹

Further light is shed upon the general long term trend of alteration in consumers' desires by study of the changes which have occurred in the quantities of goods produced and in the growth of particular industries. Recalling that the population of the United States increased approximately 50 per cent from 1899 to 1925, it is illuminating to observe that during this period the production of cigarettes increased 2,038 per cent, petroleum refining, 1,140 per cent, automobile production, 151,850 per cent, manufacture of ice, 836 per cent, canning and preserving of fruits and vegetables, 524 per cent, sugar production, 350 per cent, and the manufacture of silk goods, 322 per cent. At the same time the production of cotton goods increased only 76 per cent, boots and shoes, 48 per cent, and ship and boat building, 22 per cent. Such data as these reflect significant changes in the habits, the home lives, the clothing and food requirements of consumers. He would be a courageous but foolhardy producer who would disregard these tendencies.²

Indicators of purchasing power: Consumers' incomes. The utility of such data for the purpose of assisting the business man in predicting the future demand conditions of his product is strictly limited. They apply only to the long time movements, and on the whole they change rather slowly. As a background for the interpretation of those tendencies which are at work over shorter periods, they have value, but they do not interpret, much less predict, these shorter time movements. Changes in the volume of money or purchasing power in the hands of consumers are largely responsible for the more marked short time changes in demand.

It is however exceedingly difficult to measure accurately the total incomes of all consumers. American statisticians have worked on

¹ *Recent Economic Changes in the United States*, 1929, pp. 33-34.

² The figures in this paragraph are taken from *Recent Economic Changes in the United States*, 1929, pp. 53-54.

the problem of the measurement of national income and have produced important statistical results. For example the United States Department of Commerce, Division of Economic Research, in coöperation with the National Bureau for Economic Research, has compiled the figures presented in the table below.

INCOME PAID OUT IN THE UNITED STATES ¹

Type of payment	Billions of dollars				
	1929	1930	1931	1932	1933
Salaries, wages, and pensions	52.7	48.4	40.7	31.5	29.3
Dividends	5.9	5.8	4.3	2.8	2.1
Interest	5.5	5.6	5.4	5.3	5.2
Rents and royalties	4.4	3.7	3.1	2.4	2.3
Entrepreneurial withdrawals	13.8	12.3	9.8	7.7	7.9
Totals	82.3	75.8	63.3	49.7	46.8

Similar work has been done in Great Britain and other countries of Europe. From the standpoint of the business man these data are generally inadequate, for the reason that they are always at least a year or more behind the present, that they give total income only and not the incomes of various classes, that they are annual figures, and that they have certain statistical characteristics that detract from their usefulness as a business barometer.

There are however methods of obtaining indirectly certain measures of income changes that are serviceable to business men. The most satisfactory of these measures are the index numbers of employment and payrolls in manufacturing industries. These figures are currently computed by the United States Bureau of Labor Statistics and are adjusted for seasonal variation by the Federal Reserve Board. They are published in the *Monthly Labor Review*,

¹ Income paid out is defined as "compensation for services rendered by economic enterprises, including government units, to the investors and to persons engaged therein, which includes profits withdrawn from the business or profession by entrepreneurs." U. S. Department of Commerce, *Survey of Current Business*, January, 1935, pp. 16-18. See also U. S. Senate Document No. 124, Seventy-third Congress (1934), *National Income 1929-1932*.

For America see also publications of the National Bureau for Economic Research, W. D. King, W. R. Ingalls, and M. Copeland. For England, see the works of Professor A. Bowley and Sir J. Stamp.

the *Federal Reserve Bulletin*, and the *Survey of Current Business*. The last named gives them in greatest detail. The following are annual averages of these indexes,¹ the base (100) being the monthly average of the years 1923, 1924, and 1925.

<i>Year</i>	<i>Factory employment</i>	<i>Factory payrolls</i>
1919	107	98
1920	107	117
1921	82	76
1922	91	81
1923	104	103
1924	96	96
1925	99	101
1926	102	104
1927	100	102
1928	100	104
1929	106	110
1930	92	89
1931	78	68
1932	66	47
1933	73	50
1934	86	65
1935	91	74
1936	98	86
1937	106	102

Here are examples of these employment indexes by months:

<i>Year and month</i> <i>1938</i>	<i>Unadjusted</i>	<i>Adjusted for</i> <i>seasonal variation</i>
January	88	90
February	88	89
March	88	87
April	86	85
May	83	84
June	82	82
July	82	83
August	86	85
September	89	87
October	90	88
November	91	90
December (preliminary)	91	91

Separate monthly indexes are also published for factory employment in many individual industries.

The amplitude of fluctuation of factory workers' incomes is much greater than that of retail sales, which indicates the sensitivity and therefore the usefulness to the business man of factory workers'

¹ *Federal Reserve Bulletin*, February, 1939, p. 138.

incomes in forecasting the demand of the succeeding month or so. Of course if the more steady incomes of railroad workers, public utility employees, government workers, and salaried employees were included, the amplitude would undoubtedly be reduced.

Movement of goods from producer to consumer. At best however the data of factory workers' incomes indicate the probable changes in demand for only a short period in the future. A longer range indication of future demand is provided by a study of the flow of goods from producer to consumer. The flow of goods from the producers to the wholesale and retail markets is indicated by such measures as carloadings, exports and imports, and wholesale trade. The flow of goods from the wholesale and retail markets is indicated by such measures as department store sales and stocks on hand, mail order house sales, life insurance, and advertising. Records of these phenomena are collected and published by the United States Bureau of Foreign and Domestic Commerce, covering the period from 1919 to the present. The following is a sample portion of one of these exhibits :

MAIL ORDER HOUSE SALES ¹

(In thousands of dollars)

<i>Year and month</i>	<i>Sears, Roebuck and Co.</i>	<i>Montgomery Ward and Co.</i>	<i>Total, two companies</i>
<i>1933</i>			
January	30,620	21,840	52,460
February	30,449	21,765	52,214
March	41,071	30,797	71,868
April	44,857	37,063	81,920
May	43,463	36,150	79,613
June	43,820	35,745	79,565
July	36,316	29,075	65,392
August	39,934	32,849	72,783
September	49,167	38,556	87,722
October	53,345	46,667	100,012
November	51,215	42,295	93,510

Indexes of these barometers are published by the Federal Reserve Bank of New York in its *Monthly Review*, by the United States

¹ *Survey of Current Business*, January, 1939, p. 25.

Bureau of Foreign and Domestic Commerce in its *Survey of Current Business*, and by other agencies. These indexes are adjusted for seasonal variation, as shown in the sample below.

RETAIL TRADE INDEXES ¹

Month 1938	Department stores. Monthly average of 1926 = 100				Chain store sales (20 cos.). Average of same month in 1929-31 = 100	Monthly average of 1929-31 = 100			
	SALES		STOCKS			RURAL SALES GENERAL MDSE.		NEW PASSENGER CARS SALES	
	Unad- justed	Ad- justed	Unad- justed	Ad- justed		Unad- justed	Ad- justed	Unad- justed	Ad- justed
January	70	90	63	71	106.7	104.3	86.6	50.8	65.0
February	70	88	67	70	106.4	99.9	90.4	53.6	74.0
March	77	86	71	70	103.3	105.8	98.4	76.0	61.0
April	86	83	71	69	105.0	112.3	107.9	80.5	60.0
May	80	78	71	69	103.3	110.1	103.5	75.5	57.0
June	79	82	65	68	106.3	112.4	106.2	65.0	50.5
July	58	83	61	67	108.1	110.2	84.8	61.4	56.5
August	65	83	65	67	106.0	114.2	98.2	49.2	54.5
September	91	86	70	67	109.4	122.4	121.1	37.1	60.0
October	92	84	74	67	108.0	115.5	140.9	54.9	84.5
November	99	89	78	67	109.5	119.7	147.2	99.9	104.0

Production records. Going still further back toward the sources of income and the more sensitive indicators of consumer's demand, the business man studies the records of crop production, which reflect the prospective income of the agricultural sections. Preliminary estimates of current acreage and production are issued by the Department of Agriculture at stated intervals, and final estimates are recorded in the *Agricultural Yearbook*. The business man also scans with interest statistics showing the movement of crops. These figures are available at more frequent intervals. The table at the top of the next page is an example.

The rate of new building construction is one of the most important single influences bearing on general employment and consumer demand. Building figures are classified into residential, business, and public utility contracts. Residential building forecasts con-

¹ *Survey of Current Business*, January, 1939, p. 7.

RECEIPTS OF COTTON AT LEADING PORTS ¹

Week ending February 17

	1938 <i>Bales</i>	1939 <i>Bales</i>
Galveston	26,428	5,829
Houston	29,190	7,965
New Orleans	37,415	9,548
Mobile	3,185	233
Savannah	328	134
Charleston	1,246	8
Wilmington	534	12
Norfolk	801	305
All others	2,648	1,644
Totals	101,785	25,681
Since previous August 1	6,304,464	3,034,145

sumer demand for the commodities to equip them, etc.; business building reflects the demand for business equipment, etc.

The following is a sample of some of these records :

CONSTRUCTION CONTRACTS AWARDED ²

Figures for 37 States East of Rocky Mountains,
as reported by the F. W. Dodge Corporation

<i>Year and month</i>	<i>Value of contracts in millions of dollars</i>					
	TOTAL	RESIDENTIAL	FACTORIES	COMMERCIAL	PUBLIC WORKS AND PUBLIC UTILITIES	EDUCATIONAL
January	192.2	36.2	6.6	15.4	98.6	19.0
February	118.9	40.0	4.9	13.0	30.5	15.4
March	226.9	79.4	15.7	20.2	59.7	21.0
April	222.0	74.6	11.5	18.9	67.0	16.9
May	283.2	83.2	8.6	19.2	122.2	11.8
June	251.0	85.7	10.7	18.8	83.5	14.7
July	239.8	88.0	9.7	26.2	79.3	10.7
August	313.1	99.7	11.3	18.3	126.1	21.4
September	300.9	99.6	10.7	14.0	109.3	33.9
October	357.7	112.7	13.8	24.2	114.0	47.0
November	301.7	95.3	10.5	13.7	90.4	49.0
December	389.4	91.5	7.0	14.0	158.4	73.3

Freight car loadings, especially merchandise and miscellaneous, indicate general purchasing power and general consumption. Sepa-

¹ *Commercial and Financial Chronicle* (New York), February 18, 1939, p. 1045.

² *Federal Reserve Bulletin*, February 19, 1939, p. 145.

rate indexes are published for coal, coke, grain, livestock, forest products, ore, merchandise, and miscellaneous. The following are general averages :

FREIGHT CAR LOADINGS ¹
1923-25 av. = 100

<i>1938</i>	<i>Unadjusted</i>	<i>Adjusted</i>
January	59	65
February	57	62
March	57	60
April	55	57
May	57	58
June	58	58
July	62	61
August	63	62
September	71	64
October	75	68
November	70	69
December	64	69

Indexes of general economic conditions. Even these data however must be supplemented by indexes of general banking and credit conditions if sound judgment is to be secured. Bank debits (*i.e.*, charges by the banks against customers' accounts, chiefly for checks drawn) indicate the general volume of money transactions being consummated. These figures for centres outside New York reflect general industrial and commercial business. New York City figures are colored by speculative transactions of the stock exchanges. Postal receipts, electric power production, business failures, velocity of bank deposits, new incorporations, all play a part in providing a complete picture of future demand. The following table contains indexes of a number of these factors, of the sort that are regularly used by American business men in their attempts at forecasting demand. These indexes are computed by the Federal Reserve Bank of New York. They show for each month the magnitude of the particular factor as a percentage of "normal," a statistical concept based on past trend and seasonal variation.

Much more might be written about demand. The circumstances which determine people's purchases: custom, tradition, education, fashion, individual peculiarities, sudden whims, selfishness, al-

¹ *Federal Reserve Bulletin*, February 19, 1939, p. 138.

INDEXES OF GENERAL BUSINESS ACTIVITY ¹

	1937	1938		
	SEPTEMBER	JULY	AUGUST	SEPTEMBER
Bank debits, outside New York City	66	56	56	58 ^p
Bank debits, New York City	37	35	32	34 ^p
Velocity of demand deposits, outside New York City *	69	61	59	61
Velocity of demand deposits, New York City	45	40	36 ^r	38
Employment, manufacturing, U. S.	102	78 ^r	80	82 ^p
General price level **	161	155	154	154 ^p
Composite index of wages †	112	110	109	109 ^p
Cost of living **	153	149	148	148 ^p

^p Preliminary. ^r Revised. * 1919-1925 av. = 100. ** 1913 av. = 100; not adjusted for trend. † 1926 av. = 100; not adjusted for trend.

truism, patriotism, religion — all this furnishes an interesting field of study. Equally interesting is the subject of the creation or modification of demand, not only by business men through advertising, salesmanship, etc., but also by “reformers” through education, religion, and public appeal of various sorts, or by laws forbidding some purchases (as alcoholic liquors) and compelling others (as automobile license plates and bathing suits). On the foundation of the brief analysis of human wants already presented could be built an interesting study of how each want works itself out, in correspondence and in competition with other wants, when independent variations are considered in more than one price; but these excursions are not within the scope of this elementary treatise.

EXERCISES

1. Name various ways in which sellers may seek to distinguish their products from those of competitors.
2. Mr. A grows cotton and sells it at 10 cents per pound. Draw the appropriate demand curve for his product. If he should decide to charge a higher price, what would be the effect on his sales? If he should increase his output what would be the effect on price? Would he be justified in advertising his product? Explain, in each case.

¹ Federal Reserve Bank of New York, *Monthly Review of Credit and Business Conditions*, November 1, 1938, p. 95.

3. Mr. B is a manufacturer of toothpaste, which is attractively packaged, nationally advertised, and sold under a brand name. Last year he sold 1,000,000 tubes at 25 cents each. Draw the appropriate demand curve for his product. If he should decide to charge a higher price, what would be the effect on his sales? If he should decide to lower his price, what would be the effect on his sales? Is he justified in advertising his product? Explain, in each case.

4. In retail grocery stores the sale of "bulk" coffee, crackers, and candy has given way to the sale of similar products attractively packaged, nationally advertised, and sold under brand names. Account for this development and explain its significance.

X

THE SELLERS AND THEIR COSTS

The different viewpoints of buyers and sellers. The buyers of goods are either consumers, who seek goods to satisfy their wants, or producers,¹ who buy in order sooner or later to sell either to consumers or to other producers. In any case the goods finally reach consumers, and it is this last exchange which is the fundamental one for the study of the demand side of the forces determining price. The demand of manufacturers and dealers is not the ultimate demand; it is an indirect or reflected demand, arising from the demand of the "ultimate consumers." In seeking the causes back of demand therefore we were justified in passing over the demand of the producers in order to give our attention to the causes which determine the demand of those who seek goods to satisfy their own wants. When now we turn to the study of supply, we meet at the outset the important fact that in general there is no corresponding class of sellers who are parting with goods which might satisfy their own wants. Practically all selling is by producers, not by users.

Of course there are exceptions. Financial reverses or other changes in personal or family circumstances sometimes lead people to sell their personal effects, thus depriving themselves of their use in order to devote the money received to the satisfaction of more urgent wants. Quite similar is the case of goods which are not capable of reproduction, such as rare postage stamps, first editions of books, antique furniture, paintings by old masters, and all articles which are desired on account of some peculiar association or quality which cannot be reproduced. When any such thing is offered for sale, someone must have been prepared to give up its use in order that the purchaser may enjoy it. Even when such an article is sold by a dealer, who is himself not interested in its use,

¹ The term, producers, includes of course merchants and all others who have a part in the productive process.

we have only to go back one stage to the former user from whom the dealer acquired the article. These are obviously exceptional cases, not representative of the ordinary run of sales.

For goods in general there is thus an important difference between the ultimate buyers and the sellers. The purchaser buys a set of golf clubs in order to play the game. The merchant who sells him the clubs is not thereby giving up his opportunity to play golf, and the estimate he places upon the clubs has nothing to do with the pleasure of playing golf. The seller's conduct is not explained by a balancing of relative marginal utilities.¹

On the seller's side the motive is profit, or other pecuniary gain. This is true not only of the retail merchant from whom the consumer does most of his buying. It holds also of the wholesaler, the jobber, the manufacturer, the farmer, and all other persons engaged in the various steps of production from the extraction of the raw material to the delivery of the finished good to the final consumer. This generalization applies also to those who sell their personal services, such as laborers, lawyers, physicians, musicians, and actors. The seller makes a profit when he sells for more than his goods cost him. He avoids a loss when his selling price just equals the unit cost of his product. Cost of production thus becomes the most important consideration on the seller's side of the market, just as marginal utility is most important on the buyer's side.

Production to order or for a general market. In the early stages of the medieval town economy most of the products of manufacture were "made to order." The blacksmith, the bootmaker, the weaver did not generally begin the making of anything until there had been an order from a buyer and an agreement upon the price. Since no artisan would knowingly undertake to produce and sell an article for a price that would not cover his costs and a "fair profit," there was here a simple relation between cost and price.

¹ It is true that we must finally push our analysis of the selling side of the market back through the actual sellers to the "ultimate producers," corresponding to the ultimate consumers on the buying side. All costs thus reduce themselves to labor cost and the cost of waiting for future income, and we come finally to a principle of "marginal disutility of labor," somewhat analogous at least to the principle of marginal utility of goods. This analysis will be undertaken later, in connection with the study of wages, profits, and interest. It must wait however till we have disposed of our present problem, which has to do with the forces which immediately control the actual sellers of goods.

Present-day examples of such production are the custom tailor, the custom bootmaker, the house-building contractor, etc.

These examples do not however represent the most important type of modern production to order. In fact a very considerable part of the production of machinery, tools, structural steel, and a host of products that furnish either equipment or raw materials for other factories is made to the customer's orders. Here the relation between cost and price is not quite so direct and obvious as in the case of the simpler industries cited in the previous paragraph, owing to the complexity of the industrial processes and the consequent difficulty of knowing in advance just what the cost per unit of product is destined to be, and owing also to the different elements that may be included in cost and to the different ways of measuring cost.

The relation between cost and price becomes still more obscure when production is not to order but for a general market. Wheat, cotton, and indeed all agricultural products are obviously in this category. The same is true of most every-day articles which are sold for use to the general public rather than to other producers. Ordinary ready-made clothing, shoes, hats, bread, breakfast foods, and household furniture are a few examples from an almost unlimited list. In fact this condition is typical of the greater part of modern industry. Such goods are made up in advance of orders from buyers, and the producer takes the chance of being able later to find customers who will take his product at a profitable price. Under such circumstances we cannot say simply that sellers will not part with their goods except at prices that equal or exceed their costs. The complex relationship between cost and price will appear only after rather searching analysis.

Immediate sales problem. As the first step in our analysis of costs in their relation to price, we shall need to distinguish between what we may call *immediate* and *remote* sales considerations. The former are those that confront the seller — manufacturer or dealer — with respect to a stock of goods on hand, already produced. Such goods are in a special position, because all of the costs to which their production or acquisition has given rise have already been irrevocably incurred. The seller's sole interest in them is to realize

as large a return as possible now that they have come into his possession. Of course considerations of the more remote future will influence the selling policy which he adopts with respect to any particular lot, but the point is clear that for such goods there can be no absolute minimum below which no seller would go. Each seller would like to get back his costs and as much as possible in excess. But the cost — whatever it is — has already been paid and is generally irrelevant to the seller's decision.

The most obvious illustration occurs in the case of non-reproducible goods. The past cost of producing a rare postage stamp, a painting by a famous old master, or a rare and beautiful piece of antique furniture has nothing to do with the price at which it can be sold. In these examples the demand of buyers is such that the possessor's stock can be disposed of at prices far above cost of production. On the other hand, there are non-reproducible goods for which no one would today give as much as the cost — a painting by a past nonentity for example. To dispose of such an article the seller must accept whatever price will be agreeable to some buyer. The auction is the typical process. Cost of production has no influence on the price.

A somewhat similar situation exists in the case of articles which, though they could be reproduced, have gone out of fashion and are of so little present use that no one would consider paying as much as it would cost to make them. A search through the barns and carriage houses of the United States would disclose many an old-time sleigh. Sleighing is pretty much out of fashion, and the fact that a sleigh, still in good condition, might now cost three hundred dollars to reproduce will not enable its owner to get any such price for it. If he wants to sell he must take what he can get, regardless of cost.

Even in the case of goods which are being currently produced, the seller cannot always be sure of an immediate price equal to cost or greater. As we have seen, in a free market no seller can set both the price and the quantity he will sell. In the case of perfect competition, the seller is confronted with a market price. The wheat farmer can sell his whole stock of wheat, or any part he may choose, at the market price. He cannot exact a higher price because his

own cost of production was more than the market price, nor does he have to take a lower price because his cost was less. When competition is not perfect, each seller, as we have seen, has a price problem. If he is determined to dispose of his entire stock he will have to set a price with relation to the demand for his product, and there is nothing to guarantee that this price will equal or exceed his cost. If he sets a price equal to his cost, he may be able to sell only a part of his stock or none at all. On the other hand, an unexpected demand for some commodity by persons who will not wait for future production may enable fortunate dealers to dispose of their limited stocks at fancy prices for above cost. Raincoats and waterproof fabrics are sometimes thus sold to the crowd at a football game on a rainy day.

Reservation price. In facing his immediate sales problem, the seller does not of course have to accept a price low enough to take off his entire stock. Generally he will set a minimum which may be called his "reservation price." The application is different according to whether there is pure or imperfect competition. The wheat farmer has no occasion to have a reservation price lower than the market price, and if he sets one higher it will mean that he can sell no wheat. What the farmers frequently do is to sell part of their stock at the market price and hold out the rest for future sale. When imperfect competition prevails, setting a reservation price will generally result in only a part of the stock being sold.

Various considerations influence sellers in setting their reservation prices. In the fresh fruit and vegetable markets, as the end of the selling period approaches (say late on Saturday evening) the dealers slash their prices ruthlessly in order not to be left with goods on hand which will spoil before the next market opening. More durable goods are less subject to this influence. The cotton or wheat farmer can more safely decline to sell at the present price. Goods subject to the whims of fashion are similar to perishable goods. Thus as the summer wanes, the department stores announce "sales," which mean that reservation prices have been cut in order to avoid having unseasonable and out-of-date hats, dresses, shoes, and such goods left on their hands.

Reservation prices are strongly influenced by the seller's esti-

mates of what the future price of his product may be. The higher the expected future price, the higher will be the present reservation price. There will normally be a "spread" between expected future price and present reservation price equal to the cost of "carrying" the merchandise, consisting principally of storage cost and interest on the present value represented by the goods. For example suppose that a lumber dealer believes that the price of white oak flooring next year is likely to be \$90 a thousand feet. Assume further that it will cost him \$2.00 a thousand for storage for the coming year. If interest is at 5 per cent, he will have a reservation price of about \$83.81. If he sells at this price, his receipts plus interest of \$4.19 will amount a year hence to \$88, which is what he would then receive less \$2.00 storage cost. It will therefore be wise for this dealer to sell today as much of his stock of white oak flooring as he can dispose of at a price of \$83.81, but not to try to sell more by offering a lower price.

Sometimes a dealer finds himself in financial straits and must turn his stock into cash regardless of cost. He sells at "sacrifice" prices, for which he may have no minimum reservation at all. In times of general financial panic many dealers may find themselves in this position, and a general collapse of prices may result. On the other hand, a sudden upturn of economic prosperity or the general expectation of such upturn may simultaneously influence all dealers to raise their reservation prices and so lead to a general increase in the prices of present stocks of goods. The main importance of these considerations arises from the light they throw upon the forward looking character of sellers' decisions. Once costs of production have been incurred or once a dealer has purchased goods, these past transactions have no significance for his decisions. Of course the price that he succeeds in realizing for a stock of goods already produced may determine whether he will produce or buy any more goods of the same variety. But this leads us away from the immediate sales problem to a consideration of our next topic.

Remote sales problem: Cost of production controls. We have considered the problem of the seller in connection with the disposal of a stock of goods on hand. We saw that in some cases cost of production has no effect on price and that even in case of goods being

currently produced the price may at certain times depart materially from the cost. As regards the latter class of goods however, which of course comprises the great bulk of goods for sale, a moment's consideration will indicate that, just as soon as we pass beyond immediate considerations to the long range view, costs of production are no longer negligible but become the central factor in the seller's problem.

Cost and the lower limit to price. Business men, farmers, and producers generally will not be able or willing to remain in business indefinitely if they sell their products at prices which will not cover their costs. However much a business man might like to be able to lower the price of what he sells in order to expand his business, the necessity for covering costs will prevent him from carrying this procedure beyond a certain point. Thus in a general sense, we may say that costs set lower limits to prices.

However even this simplified statement is a dangerous one, since there are few words in use by economists and business men which are more highly technical in their practical applications than the word *cost*. Economists, accountants, national and state commissions spend much of their energy in trying to decide exactly what items should be included in the cost of production and what cost concepts are significant for a particular situation. It is not necessary here to enter into these problems in detail, but we must note at the outset that costs include, not only the payments made by the entrepreneur to other persons, but also interest upon his own capital and an adequate reward for his own labor of management. The owner of capital has the choice of employing it in business or lending it to someone else. In the latter case he receives interest, and he will therefore be unwilling to employ it himself unless he can expect an equal return from his own business. In like manner the entrepreneur generally has before him the two alternatives of continuing to run his own business or of hiring out to an employer. His choice will ordinarily depend upon the relation between the wages or salary which he could obtain as a hired worker and the income that accrues to him from his business. Unless the latter is in the long run at least as great as the wages or salary he could command, the business man will ordinarily

not continue as an independent producer. Production cannot be carried on under our present economic system without the services of the business manager. The minimum reward required to induce him to perform his function is therefore just as much a necessary cost of production as the wages of laborers or any other cost. This minimum is sometimes called "wages of management," to indicate that this part of the income of a business man is only what he could presumably earn as salary or wages if he should choose to work for an employer instead of running his own business.

In this and in the following chapters we shall study with considerable care the relationship between costs and prices, both in an individual firm and in an industry. Until we have investigated this relationship further, all that we can say is that, in a loose sense, cost of production is the factor which prevents prices from falling below certain limits or remaining long below certain levels even when competition is severe.

The upper limit to price. Of course producers would gladly receive and will generally seek to obtain prices which are above any minimum set by costs of production. What are the factors, if any, which set an upper limit to prices? Again it is impossible, until we have studied the problem in considerable detail, to answer this question finally, but just as it is clear in a general way that costs prevent prices from falling to zero, it may be seen that costs in any business or industry tend generally to prevent prices from being raised without limit.

Let us assume that several competing bakeries are offering bread to the consumers of a certain community. Each baker is striving to make his profits as large as possible, which he may accomplish either by a high price or by large sales. If in choosing the first of these alternatives, he sets a price which is substantially greater than the costs of production among competing bakeries, he will set a margin within which rival bakeries can underbid him while still covering their costs. While there may be a few customers who prefer the bread of this particular baker and will continue to buy it, the majority of his customers will tend to leave him in order to buy their bread more cheaply. It is thus that the costs of his competitors, or the general or "bulk-line" level of costs in the

industry, tend, through their effect on the demand for any particular baker's bread, to prevent him from setting his own price too high. The more elastic the demand for a particular seller's product, the more will he be limited in his power to raise prices by the costs of production in the industry.

Prices are controlled, not by costs of existing goods, but by future costs. The cost of production which is one of the important factors in price determination is not past cost already incurred. Suppose a new invention has greatly reduced the cost of making a certain product. Producers and dealers will generally find it impossible to dispose of all their goods on hand at a price equal to what it cost to produce the existing stocks. The fact that the stock may now be replenished at a lower cost will lead competing producers to offer the product at a price related to the new cost of production. Those who are stocked up will find themselves compelled to offer their stocks at the new price. On the other hand, consider the case of a change which increases the cost of a certain commodity, such as a rise in freight rates or an agreement fixing a higher scale of wages. There will now be a new and higher cost scale governing all future production. Those who have stocks of the commodity on hand, knowing that they cannot be replenished except at a greater cost, will not be likely to go on offering the commodity for sale at a price just covering the old cost of production. They will find that they can get a price governed by the new cost even for those articles which were produced at the former lower cost, and this will control the supply offered by them.

In normal times the prices at which producers offer goods are thus controlled by future costs of reproduction and little, if at all, by the past costs of existing stocks. Of course no wise producer closes his eyes to the past or present scale of costs, and if there appears no reason to anticipate a change in the near future, selling policy is apparently determined on the basis of past costs. But that is simply because costs already experienced have turned out to be the same as the later sales price, and it is only as an indication of what future costs may be expected to be that past or present costs enter into the problem. In making offers for future delivery, business men are governed by their estimates of future conditions

of production and future costs, and the selling prices at which they thus arrive will generally determine also the prices at which existing stocks will be offered.

In spite of the logic of this conclusion, the reader may possibly find it hard to give up the notion that manufacturers and dealers, through their systems of cost accounting, do fix the prices of their products on the basis of what it cost to produce them plus a pre-determined rate of profit. Cost accounting does, it is true, have its effect, particularly when, cost conditions being for the time comparatively stable, there appears no reason to expect that future costs will be materially different from those of the immediate past. But the final authority of future costs is not overthrown, as a careful analysis of the "load" added for profits will often disclose. No system of cost accounting will long prevent producers from taking advantage of changed conditions which enable them to make more than the usual or expected profit on existing stocks, as they did for example in the boom period of rising prices that culminated in 1920. And by the same token producers cannot, by the simple device of fixing their selling prices on the basis of "cost plus," avoid selling existing stocks at a loss when changed conditions lower the scale of future costs.

The lingering influence of past costs. Though prices are not fixed by past costs of production, in abnormal times these costs do have a retarding effect on the movement of prices. When the whole economic structure is upset by a severe depression and sellers are overstocked and the manufacture of new goods declines, there is a general disinclination "to take a loss," and a general ignorance of the fact that losses will have to be taken. Sellers vainly hope for a quick revival of business, which will permit them to recoup their old costs of production. Ultimately under these circumstances the prices of surplus goods are determined, not by cost of production, but by the ability and willingness of buyers to take them off the market. Before the depression gets to this stage however the downward movement of prices is retarded by the reluctance of sellers to let them go below their old costs of production any more than they think they have to. In the case of minor recessions also this is a factor of considerable influence.

Analysis of cost: Fixed and variable costs. Our next task is to obtain a definite idea of the meaning of unit cost of production. For this a careful analysis of cost is required. Let us suppose that a certain manufacturer is producing shoes, using a plant (land, factory, machinery, etc.) which cost \$1,000,000. The interest on this capital is of course one of the elements in the cost of producing shoes. It is obvious that, this sum having once been invested in the plant, the interest will go on indefinitely, without regard to the number of shoes produced. Indeed the interest will continue though no shoes at all are produced; it cannot be escaped even by closing down the factory. Even if this particular owner should escape by selling the plant, the purchaser would have to assume the burden. Only by going into bankruptcy can the enterprise shuffle off its contractual interest obligation, and then the creditors suffer a lump sum loss equal to the future interest payments.

There is another group of costs whose amount has little dependence upon the immediate number of shoes produced; this group includes such costs as the salaries of the principal officers and department managers of the factory, the payments for insurance, for lighting and heating the factory, etc. These costs are unlike the interest on capital however in that they would generally cease if the plant were closed down permanently.

Finally there are costs which are related closely to the quantity of the product, so that their amount varies with the amount produced. Thus the cost of leather and other materials, wages, and certain other costs will vary according to the number of shoes produced by the factory.

Production costs may thus be classified as follows:

1. Fixed costs: (a) interest, (b) other fixed costs
2. Variable costs.

The fixed costs are those whose amount does not vary with the quantity of goods produced; they are subdivided into (a) *interest* on the capital invested, which is permanently sunk in the enterprise and must be borne indefinitely (or until bankruptcy) even though the business should be shut down, and (b) *other fixed costs*, which would not have to be incurred if the business were given up,

but whose amount has little or no relation to immediate output.¹ *Variable costs* are those whose amount varies with the amount of goods produced, although not necessarily in the same proportion. In the long run, all costs tend to become variable, for, though certain costs may be fixed within a given range of production, they also are subject to variation if production is permanently increased or decreased by very large amounts.

A numerical example will help us to arrive at the relation between these several groups of costs and the total cost of production per unit of product. Returning to our example of a shoe factory with \$1,000,000 of capital invested in plant, let us assume that the annual interest charge, at 5 per cent, is \$50,000. Let us assume further that the other fixed costs are \$450,000 a year and that the variable costs are \$3 per pair of shoes produced. What will be the total cost of making a pair of shoes? Evidently it will depend upon the number of pairs produced. The following table shows the results of three separate assumptions as to the output:

COST OF PRODUCING A PAIR OF SHOES

	<i>First case</i>	<i>Second case</i>	<i>Third case</i>
Amount of product	500,000 prs.	1,000,000 prs.	2,000,000 prs.
Interest	\$ 50,000	\$ 50,000	\$ 50,000
Other fixed costs	450,000	450,000	450,000
Variable costs (\$3 a pair)	<u>1,500,000</u>	<u>3,000,000</u>	<u>6,000,000</u>
Total costs	\$2,000,000	\$3,500,000	\$6,500,000
Average cost per pair	\$4.00	\$3.50	\$3.25

It should be especially noted here that this example does not represent a historical sequence; it sets forth rather the results of three simultaneous possibilities as to the quantity that might be produced. The illustration brings out with clearness the principle that, under these assumptions, the average unit cost of production varies inversely with the amount produced. The reason is simple enough; *i.e.*, since the interest and other fixed costs are constant in amount, the part of such costs assigned to each unit becomes successively less as the number of units increases. There is within limits a general tendency for the unit cost of production of each producer to decrease as he is able to increase his product.

¹ The term "overhead costs," or simply "overhead," is frequently used to refer to what we have called fixed costs.

As a practical example of this principle, the following incident is pertinent :

"A substantial reduction in the contract price of newsprint paper is to be announced shortly by the International Paper Co. This reduction is a result of the International having secured a contract from the Hearst interests for a large tonnage to be delivered over a period of several years at a price below the present price. This may stabilize the newsprint market, which has been threatened with collapse for some time because of the failure of consumption to keep pace with the heavy increase in production capacity. The general impression is that the new lower price shortly to be announced by International will so reduce the profit margin for other producers that further attempts to erect additional newsprint machines will be discouraged. The mills are now operating at about 80% capacity."¹ The effect of the Hearst contract on the International Paper Company's cost of production was thus: Former price, \$60.00 per ton; production, 600,000 tons per year. New price, \$52.00 per ton; production, 750,000 tons per year.

Large scale savings in variable costs. In the previous illustrations we see unit cost lowered by a "spreading of the overhead." By this we mean a reduction in average cost through the application of fixed costs to a greater volume of production. This is very important, but it is not the only way in which unit cost may be reduced by increased output. Buying raw materials in larger quantities often makes it possible to procure them at lower prices. Thus an important part of the variable cost embodied in each unit of product is reduced. Larger production often makes possible greater economy in the use of raw materials and permits the manufacture of by-products which reduce variable unit costs. In similar manner, freight rates and other shipping charges, if included in total cost of a unit of product, may be reduced by an increase in total production. Within certain limits increase in output also makes possible greater economy in the use of labor—an exceedingly important item which may be generally classified as a variable cost.

But there are very definite limits to all these tendencies. Even if one million pairs of shoes can be produced more cheaply per pair

¹ *Barron's*, Nov. 5, 1928, p. 11.

than half a million, and two million pairs more cheaply than one million, it does not follow that there would be further reduction for three, four, ten, twenty million pairs. Let us return to our numerical illustration, in which it was assumed for simplicity's sake that variable costs were \$3.00 a pair. Evidently when half a million pairs were made, the plant and the services for which fixed costs were incurred were not being fully utilized; otherwise it would not have been possible to increase the output to a million pairs and then to two millions with the same plant and the same general services. Sooner or later if production is continuously increased the time when plant and general services are fully utilized approaches.

Expanding output and unit cost. If the firm wishes to increase the rate of output beyond this point, it will have to make a choice. Usually it will be possible to increase output further by resorting to various devices which permit still more intensive utilization of the existing plant. If the firm is already running two shifts, a third shift can be added, or the plant may be kept open on Saturdays and Sundays and holidays. Sometimes the actual rates at which operations are performed can be speeded up by increasing substantially the labor force. Still another possibility is that production may be increased by making use of old equipment, less efficient than that ordinarily employed, which is kept to be used under exceptional circumstances. Because these ways of increasing output are likely to be open to a producer, it is usually inaccurate to speak of a plant as "operating at capacity." Nevertheless if output is increased in this way, it is almost certain to involve a large increase in variable unit costs. If men are hired to work overtime they must usually be paid at a higher rate. As the labor force is increased beyond its normal size, management and supervisory officers are overloaded and function less efficiently. As marginal equipment is called into use, variable costs will be increased disproportionately because of the inferior efficiency of the equipment. In some cases a factory which had been previously shut down may have to be reopened, and here again the resort to less modern equipment will push up variable costs.

The alternative choice before the business man who seeks an increase in output is to make additional investment in plant and

an increase in the staff of executive officers, etc. This will mean an increase in interest and other fixed costs, and so in the total cost per unit. When the capital investment and the executive staff are increased, the increase will at the start almost certainly be greater than sufficient to care for the increased product which can be immediately sold. Therefore there will again be a situation of unutilized plant, with the possibility of another succession of reductions in unit cost as output is further increased.

Usually the producer will choose the first of these two methods of increasing his output if the increase is in response to what he believes to be a temporary increase in demand. For instance during the World War, when it was obvious that the demand for many products was abnormal or when the construction costs for new plants were prohibitively high, much old machinery and many old factories were called into use. The variable costs in these plants were so high that it was unprofitable to use them except as increased demand for their products made it possible for them to sell their entire output at very high prices. On the other hand, if the increase in output is expected to be permanent, as when the demand appears to be growing steadily, the firm will usually choose to expand its plant and its overhead and thus incur larger fixed costs.

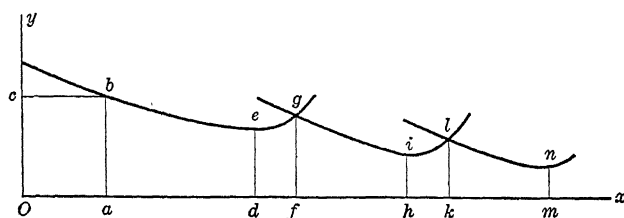


FIG. 12

Both possibilities may be illustrated by a diagram as in Fig. 12. The vertical lines such as ab measure the average cost of producing one unit when the number of units produced is measured by the corresponding horizontal lines as cb . If when the quantity reaches Od the firm chooses to expand output by a more intensive utilization of existing plant, average costs begin to rise rather sharply, as illustrated by the upward slope of the curve beyond e . If, on the

other hand, the firm decides to obtain the additional output by increasing its plant and so its fixed costs, its position will be different, as illustrated by the new cost curve passing through the point g .

It will pay the firm to use the old plant more intensively until the output, Of , has been reached. Any output larger than Of may be produced with lower unit costs in the expanded plant than by intensive utilization of the old plant. If the producer expects his output on the average to be larger than Of in future, it will pay him to expand his plant. The same situation will arise when output has grown up to the level Ok . It will again pay the producer to expand his plant if he expects his future output to be permanently larger than this.

The limit to large-scale enterprise. Because of the advantages of an increased scale of production it is likely that after each enlargement of the enterprise the unit cost of production will decline to a lower point than was reached when the plant was previously fully utilized. Returning to the diagram, hi is likely to be less than de . But this will not ordinarily go on indefinitely, since there is in most lines at least a limit to the size of the business unit, beyond which efficiency declines. In large enterprises, this limit may frequently be reached while the physical equipment is still not completely utilized. The laboring force may become too large and unwieldy, and wage costs may be rising, and managerial efficiency may decline through inability to cope with the problems of an overgrown business. An enormous enterprise like the United States Steel Corporation suffers in comparison with its smaller rivals from many disadvantages which have much the same effect as increasing costs. Such a large enterprise is apt to be unwieldy and bureaucratic. It is often difficult for decisions to be made quickly. The press of routine business may be so heavy that the executives will have little time to supervise costs with care and to attend personally to the details of the firm's operations. Referring again to our example, we may assume that a business unit sufficient to produce the quantity Om is the maximum size for efficiency. When this plant is fully utilized, further increases in plant will not permit of lower unit costs. Graphically, the successive cost curves for larger plants will tend to be, not lower, but higher, than before.

When all the producers in any industry are in this state; *i.e.*, with enterprises of the size of maximum efficiency and plants generally fully utilized, the limit to the advantages of large-scale production has been reached.

Individual costs dependent on general conditions. In the foregoing discussion our attention has generally been directed to cost conditions in individual concerns. This should not blind us to the fact that the costs which each producer has to consider are affected by general conditions in the particular industry and even by conditions in all industries together. Among the most important of these conditions, we may note the general state of technology in the industry, the general supply of raw materials needed in the industry, the general supply of the particular types of labor needed by the industry, the development of transportation facilities as needed by the industry, and the development of marketing facilities needed by the industry. Individual cost depends, not only on the volume of operations and investment in the individual firm, but also on the volume of operations and investment in the entire industry and in industry generally. Taking account of these more general influences, whether in theory or in practice, is a difficult matter, since we can never have the same degree of exactness or completeness of knowledge of general conditions as we do of the immediate conditions within a particular plant.

This relationship between particular and general cost conditions should be kept in mind equally in connection with the discussion of average and marginal costs which occupies the remainder of this chapter.

Marginal cost: In general. Analysis of the way in which cost of production enters into producers' decisions will be facilitated by the introduction of a new concept; namely, *marginal cost*. Hitherto we have confined our attention to the producer's average cost, and it is clear that if a producer wishes to know what profit he is making or what loss he is suffering at any particular time, it is a comparison of price and average cost that will give him his answer. However economic analysis is concerned with the decisions which producers make and the price and production policies which they adopt. We must therefore investigate cost concepts which will

enable the producer, not merely to assess his position at any given time, but also to judge the profitability of possible changes in policy. A firm may be in a very satisfactory position, but if its manager is energetic, he will constantly have in mind the possibility of changing his price or expanding his output by some other means.

Suppose that he is considering a change in policy which would lead to an increase in his sales; how can he determine whether or not the change in policy should be made? Clearly he will have sooner or later to make a comparison between two significant quantities. The change will presumably affect the revenues of his firm in some way, and it will also affect the firm's costs. If it will reduce costs more than it reduces revenues or increase costs less than it increases revenues, then the change will be a profitable one; otherwise not. Decisions of this sort we shall analyze in more detail in subsequent chapters, but it is important for cost analysis to note that the executive in our example will wish to know by how much his cost would be increased or decreased if the proposed change in policy were carried through.

The new cost concept which we shall now consider is designed to be of assistance when such decisions as these are made. We may define *marginal cost* as *the amount per unit by which a small change in output changes total cost*.¹ The significance of the concept may perhaps be made clearer if we think of marginal cost as the difference in total cost that will be occasioned by a change of one unit in output. This description makes marginal cost exactly parallel with the concept of marginal utility as we have defined it. It is obvious however that it is generally impossible to measure the increase in cost that will result if just one more unit is produced or the saving in cost if one less unit is turned out. For instance if a shoe factory had been planning to turn out 500,000 pairs of shoes, and we attempted to discover by how much its cost would be increased if it turned out 500,001 pairs, the calculation would be impossible. In actual practice a fairly rough measure is generally the best that can

¹ Since this amount would be calculated by dividing the increase in cost by the number of units added to output, marginal cost is a *rate*, and it is sometimes defined as "the rate at which a small change in output changes total cost."

be obtained. However the more minutely changes in total cost can be calculated, the more accurate is any measure of marginal cost.¹ As a working concept we define marginal cost as the amount per unit by which total cost changes with a small change in output; and by a small change in output we mean the smallest change for which the effect on cost can be calculated. Small and large are always relative terms. In practice the smallest change upon which marginal cost can be calculated may be so large as to necessitate the erection of a new wing in the factory, or even an entire new plant.²

Marginal cost and average cost. The nature of marginal cost and its relation to average costs may be given more concrete illustration by returning to the example of the shoe factory which has already served us. Let us revise the assumptions however in two respects; in the first place, let us consider this the business unit of most efficient size, and second let us recognize the probability that variable costs will not be \$3.00 a pair throughout.

COST OF PRODUCING A PAIR OF SHOES

Case 1

Amount of product	490,000 pairs	500,000 pairs
Interest	\$ 50,000	\$ 50,000
Other fixed costs	450,000	450,000
Variable costs	<u>1,967,000</u>	<u>2,000,000</u>
Total cost	<u>\$2,467,000</u>	<u>\$2,500,000</u>
Change in total cost		\$33,000
Change in output		10,000 pairs
Marginal cost		$\$33,000 \div 10,000 = \3.30
Average of variable costs		$\$2,000,000 \div 500,000 = \4.00
Average of all costs		$\$2,500,000 \div 500,000 = \5.00

In this illustration it is assumed that 10,000 pairs of shoes is the smallest change in contemplated output upon which we can base a calculation of marginal cost. The effect in Case 1 above is to change total costs by \$33,000. Dividing this amount by 10,000, we

¹ The pure mathematical concept of marginal cost is based on the assumption of infinitesimal changes in cost and output.

² Cost accountants have developed a working concept known as *differential cost* closely related to the economist's idea of *marginal cost*.

have a fairly accurate measure of marginal cost. \$3.30 is the rate per pair at which total cost changes with an increase in production from 490,000 to 500,000 pairs, or with a decrease in production from 500,000 to 490,000 pairs. \$3.30 is said to be the marginal cost of 500,000 pairs of shoes. According to our general definition it would be logical to call \$3.30 the marginal cost of either 490,000 pairs or 500,000 pairs; but to prevent confusion it is generally understood that, when a small change in output takes place through addition, the rate of additional cost is to be taken as the marginal cost of the quantity which includes the added units. Thus the same marginal cost is not attributed to two different quantities; and marginal cost through addition and through subtraction are regarded as the same for any particular quantity.¹

It is important to note that only in special situations is either the average of variable cost or the average of all costs equal to marginal cost. In Case 1 above, marginal cost (\$3.30) is less than the average of variable costs (\$4.00), which incidentally has to be less than the average of all costs (\$5.00). The relative level of marginal cost indicates that the rate of increase in cost is less at this point than lower down the scale of production. Otherwise marginal cost could not be less than the average of all variable costs up to this point. We turn now to other possibilities regarding marginal cost and average cost in this same shoe factory:

COST OF PRODUCING A PAIR OF SHOES

Case 2

Amount of product	990,000 pairs	1,000,000 pairs
Interest	\$ 50,000	\$ 50,000
Other fixed costs	450,000	450,000
Variable costs	3,473,000	3,500,000
Total cost	<u>\$3,973,000</u>	<u>\$4,000,000</u>
Change in total cost	\$27,000	
Change in output	10,000 pairs	
Marginal cost	$\$27,000 \div 10,000 = \2.70	
Average of variable costs	$\$3,500,000 \div 1,000,000 = \3.50	
Average of all costs	$\$4,000,000 \div 1,000,000 = \4.00	

¹ Compare the similar convention with regard to marginal utility, Chapter VII.

COST OF PRODUCING A PAIR OF SHOES (*Continued*)*Case 3*

Amount of product	1,990,000 pairs	2,000,000 pairs
Interest	\$ 50,000	\$ 50,000
Other fixed costs	450,000	450,000
Variable costs	5,678,000	5,700,000
Total cost	\$6,178,000	\$6,200,000
Change in total cost	\$22,000	
Change in output	10,000 pairs	
Marginal cost	$\$22,000 \div 10,000 = \$2\ 20$	
Average of variable costs	$\$5,700,000 \div 2,000,000 = \2.85	
Average of all costs	$\$6,200,000 \div 2,000,000 = \3.10	

Case 4

Amount of product	2,490,000 pairs	2,500,000 pairs
Interest	\$ 50,000	\$ 50,000
Other fixed costs	450,000	450,000
Variable costs	6,970,000	7,000,000
Total cost	\$7,470,000	\$7,500,000
Change in total cost	\$30,000	
Change in output	10,000 pairs	
Marginal cost	$\$30,000 \div 10,000 = \3.00	
Average of variable costs	$\$7,000,000 \div 2,500,000 = \2.80	
Average of all costs	$\$7,500,000 \div 2,500,000 = \3.00	

Case 5

Amount of product	2,990,000 pairs	3,000,000 pairs
Interest	\$ 50,000	\$ 50,000
Other fixed costs	450,000	450,000
Variable costs	8,950,000	9,000,000
Total cost	\$9,450,000	\$9,500,000
Change in total cost	\$50,000	
Change in output	10,000 pairs	
Marginal cost	$\$50,000 \div 10,000 = \5.00	
Average of variable costs	$\$9,000,000 \div 3,000,000 = \3.00	
Average of all costs	$\$9,500,000 \div 3,000,000 = \$3.16\frac{2}{3}$	

In all these cases the change in total cost is equal to the change in variable costs. This is because interest and other fixed costs have been assumed constant throughout this wide range of production. The assumption is obviously artificial, but it simplifies the analysis. If we were to consider changes in interest and other fixed costs, these changes would enter into the determination of the various marginal costs. Fixed costs, not being constant throughout, are

themselves subject to change at certain points in the possible range of production.

Cost schedules and cost curves. The following represents a recapitulation of our findings in the five previous cases :

SCHEDULE OF UNIT COSTS IN THE PRODUCTION OF SHOES

	<i>Possible outputs</i>	<i>Averages of variable costs</i>	<i>Averages of all costs</i>	<i>Marginal costs</i>
Case 1	500,000 pairs	\$4.00	\$5.00	\$3.30
Case 2	1,000,000 pairs	3.50	4.00	2.70
Case 3	2,000,000 pairs	2.85	3.10	2.20
Case 4	2,500,000 pairs	2.80	3.00	3.00
Case 5	3,000,000 pairs	3.00	3.16 $\frac{2}{3}$	5.00

This schedule presents a series of simultaneous possibilities in the sense that, beginning with a given moment, any of the indicated possibilities might be realized at the expiration of the period necessary for production. If the output column and any of the cost columns are taken separately, these two columns constitute in themselves a specific kind of cost schedule. For example the output column and the marginal cost column constitute a marginal cost schedule. In the combined cost schedule as it stands, all of the figures in any given line are definitely related to each other.

By the same general methods previously used in constructing utility and demand curves, the schedule of costs can be converted into graphical form. Since the table includes however three different kinds of unit costs, its graphical representation is not one curve but three, as shown in Figure 13. The curves, cc' , vv' , and mm' , respectively represent various possibilities as to the average of all costs, the average of variable costs, and the marginal costs, under the given circumstances. These curves really show more than the cost schedule. Many intermediate cases are understood to be included, and all three curves have been smoothed. If we were to consider changes in interest and other fixed charges, definite breaks (discontinuities), such as appear in Figure 12, would also have to appear in the curves in Figure 13.

Let us consider the relation between these curves and our numerical illustrations. Points b , d , and e refer to Case 1. The distance Oa along the horizontal axis or base line corresponds with 500,000 pairs

of shoes. Height ab indicates the average of all costs for this quantity; height ad represents the corresponding average of variable costs; and ae , the marginal cost. We note that at points b , d , and e respectively all three curves decline downward to the right.

Our findings in Case 2 are marked by points h , i , and g . The distance Of corresponds with 1,000,000 pairs of shoes. With production at this point, the average of all costs and the average of variable

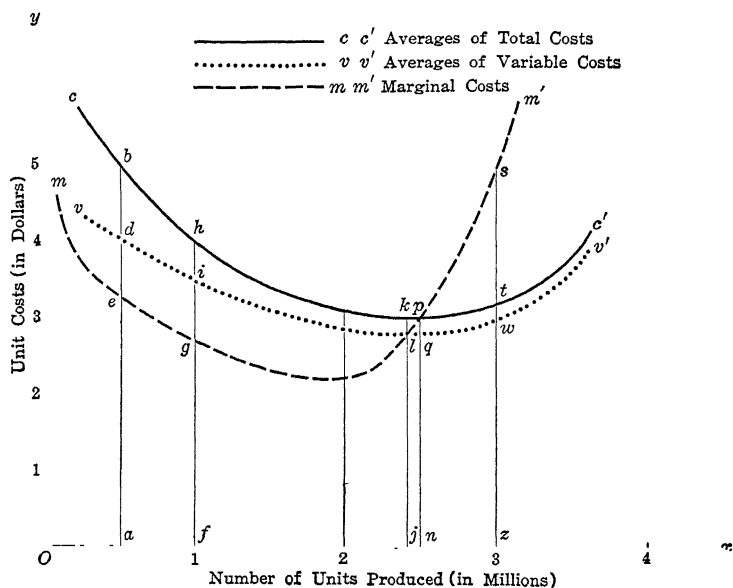


FIG. 13

costs and the marginal cost are all lower than in the previous case, and they will continue to decline with a further increase in production.

The perpendicular line without any letters refers to Case 3. We note here that marginal cost has reached its lowest level, but the two averages of cost are still declining. The increase in marginal cost beyond Case 3 checks the decline in the averages, but it will not raise them for some time yet.

Points k and l refer to an intermediate case between Cases 3 and 4. The point l is an intersection between the marginal cost curve and the average variable cost curve. Since l lies on both these

curves, the marginal cost of quantity Oj and the average of the variable costs of this quantity are both represented by the height $j\bar{l}$. In other words, the marginal cost and the average of the variable costs of this quantity are equal. In this intermediate case the average of variable costs is at its lowest level. A marginal cost curve always intersects a corresponding average variable cost curve at its lowest point. Beginning at the point l , the average variable cost curve rises under the influence of rising marginal costs. Since the average of variable costs however is not entirely governed by marginal cost, vv' does not rise so steeply as mm' .

The proof of these relations is as follows. Whenever the marginal cost is less than the average of variable costs, it is evident that the addition of another unit will increase total cost by an amount less than the average up to that point; therefore the average will be decreased. This is the case of all positions to the left of the point l , and in that area therefore the curve of average variable cost is descending. On the other hand, when marginal cost is greater than average variable cost, an additional unit will increase the average. This is true of all positions to the right of l , and in this area the average cost curve is therefore rising. Only at point l , where marginal cost is the same as average cost, is the average cost curve neither descending nor ascending; l is therefore the lowest point on the curve. A similar proposition, similarly demonstrated, obviously holds of the intersection of the marginal cost and average total cost curves, which intersect at point p .¹

Points p and q refer to Case 4. Here marginal cost and the average of all costs are the same. p is approximately the lowest point on cc' . The intersection between the marginal cost curve and the average total cost curve, as shown in Figure 13, is typical of the general way such curves always intersect.¹ This intersection is far more important than the one previously noted. The quantity On is the quantity which can be produced at lowest possible unit cost, including both fixed and variable charges. The point p is what we have called the point of full utilization, where efficiency is at the

¹ If marginal cost is calculated on the basis of infinitesimal changes, the marginal cost curve intersects both average cost curves at their respective lowest points exactly. With larger increments, the intersections are approximately at the lowest points.

maximum as indicated by the minimum average cost. Beyond that point all unit costs increase because the enterprise is being pushed beyond the point of maximum efficiency.

Such a contingency is represented by points s , t , and w , corresponding to Case 5. Case 5 is more important than it first seems, for, when prices are sufficiently favorable and productive facilities are limited, the quest for profits drives an enterprise into just this sort of situation.

EXERCISES

1. Assume the following conditions of cost for the production of shoes in the X Shoe Company plant:

<i>Output</i>	<i>Fixed cost</i>	<i>Variable cost</i>	<i>Total cost</i>
10,000	\$100,000	\$ 31,000	\$131,000
20,000	100,000	60,000	160,000
30,000	100,000	90,500	190,500
40,000	100,000	123,000	223,000
50,000	100,000	158,000	258,000
60,000	100,000	196,000	296,000
70,000	100,000	239,000	339,000
80,000	100,000	289,000	389,000
90,000	100,000	350,000	450,000
100,000	100,000	429,000	529,000

- (a) Prepare a cost schedule showing for each output the average fixed cost, average variable cost, average total cost, and marginal cost of producing that output.
- (b) Plot on a sheet of graph paper the corresponding cost curves.
- (c) Approximately what is the optimum size, or point of maximum efficiency, for this concern?

XI

COST OF PRODUCTION AND PRODUCTION POLICY

In the chapter immediately preceding we have inquired into the nature of production costs and have become acquainted with the relations between total costs, variable costs, and marginal costs. We may now take the next step toward learning the general relationship between cost of production and price, by investigating the way in which producers' costs will determine their policies as to the quantities they will produce or the prices they will charge.

Under pure competition : Price above average cost. Let us start with the assumption of a condition of pure competition, under which, as we have learned, the producer has complete freedom of choice as to his output in so far as the capacity of his plant permits; he must however accept the market price. Having no price problem, the important question confronting him is how much to produce for sale.

In order to be specific, we may return to the shoe factory of the last chapter, positing now a state of pure competition. Referring again to Figure 13, let us assume that the price is equal to zs (\$5.00). From the marginal cost curve, it is evident that the producer can afford to sell the number of units represented by Oz (3,000,000 units). The price zs (\$5.00) not only covers zw , the average of variable costs, but also covers zt , the average of all costs (\$3.17) and permits a clear profit of \$1.83 a pair. What is much more important to note however is that when the producer is making Oz (3,000,000) units, the price is just equal to his marginal cost (also represented by zs). This means that the price covers the additional cost incurred by an increase in production from 2,990,000 to 3,000,000 pairs; and that it more than covers increases from smaller amounts, as indicated by the marginal costs of quantities less than Oz . If the producer should stop short of Oz units, he would

be in position to increase his profit by expanding output, for any increase in output up to Oz would add less to his costs than it would to his total receipts. Each additional unit that he sells will bring in \$5.00, the ruling price. Each additional unit that he produces until he is turning out Oz units will increase his total costs by less than \$5.00. Thus it would clearly not pay him to stop short of this output.

These conclusions are illustrated graphically in Figure 14, which is based on a simplified version of Figure 13. When selling the quantity Oz at price zs , the average profit per unit is represented by ts ,

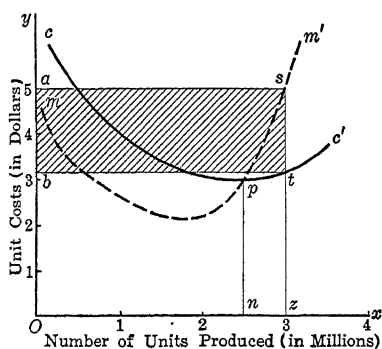


FIG. 14

being the difference between the price and the average cost for this quantity. The total profit is represented by the shaded rectangle $abts$, which is the product of the quantity by the average profit per unit and at the same time is the difference between the rectangle representing the total receipts, $Ozsa$, and the total cost rectangle, $Ozbt$.

The same reasoning demonstrates that the producer will not be able to sell more than Oz units without lowering his profit. From the marginal cost curve it is clear that if he sells 3,010,000 units instead of 3,000,000 units, the additional 10,000 units will increase his total costs at a rate higher than zs (\$5.00), whereas his receipts will increase at exactly \$5.00 a pair. Therefore it does not pay the producer to sell an extra 10,000 units at price \$5.00, and certainly no larger amount. If he does, he may still have a profit on his operations as a whole, but he will be making his total profit less than it need be. The amount which this manufacturer will produce and sell at price zs (\$5.00); that is, the output which will render his profits a maximum, is the quantity Oz (3,000,000 units). By reference to the rising portion of the marginal cost curve we can determine the amount the producer will sell at any other price, provided this price is greater than np . The reader will note that, whenever

the price is above np , the shoe factory will produce an output greater than On and will always be making a profit, since to the right of On the marginal cost curve lies above the average cost curve.

Price below average total cost and above average variable cost. But what will be its situation if the price falls below np ? As an aid to the analysis of this question, Figure 15 is presented, being an

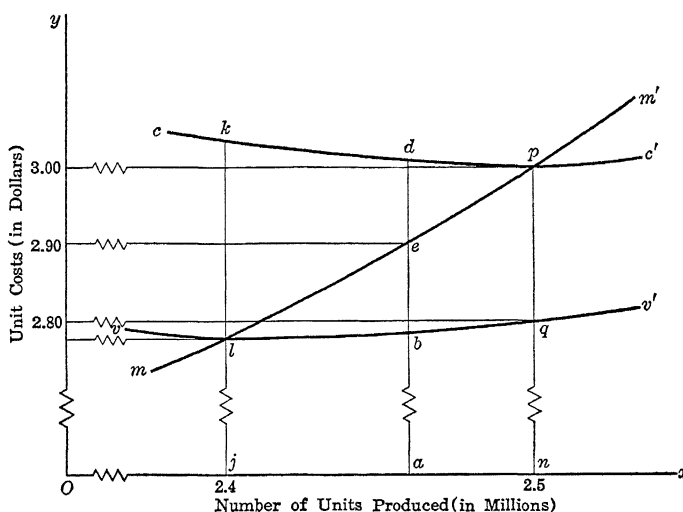


FIG. 15

enlargement of the pertinent part of Figure 13.¹ Clearly the firm will lose money, no matter what quantity it decides to produce, because np is the lowest point on its average total cost curve. If the price is less than np , there will be no output which would so reduce the firm's costs that it could produce and sell that output at a profit. Suppose our shoe factory finds itself in just this position, with a price, let us say, of \$2.90 a pair. No amount can be produced in this factory and sold at this price without a loss. But if the quantity Oj could be produced and sold for a price as high as $j\bar{l}$ (\$2.78), the producer would be fully reimbursed for his variable costs, though not recovering his fixed costs. At a price of \$2.90

¹ In this one case we have felt justified in violating one of the rules of graphical presentation by presenting broken scales on both axes.

this manufacturer can do better than that. He can cover all his variable costs and have a little toward paying fixed costs. This, though still leaving him "in the red," is better than closing the factory, since in that case his fixed charges would still go on, with no income at all to pay them. If in the present example the shoe manufacturer decides to continue in business, the quantity he will produce is indicated by the segment lp of the marginal cost curve. He will produce such quantity, greater than Oj , as will make his marginal cost equal to the price (\$2.90 according to our assumption). This is shown graphically by the point e , where the marginal cost curve intersects the average variable cost curve. This manufacturer will produce Oa units; his average variable cost is measured by ab (about \$2.79), and the selling price by ae (\$2.90). He is obviously incurring an average loss per unit measured by the line ed , the difference between his average total cost and the price. But he will be receiving the greatest possible sum in excess of his variable costs. In other words he will minimize his losses, for the time being.

Price below average variable cost. Finally if the price of shoes should drop below jl (\$2.80), the shoe manufacturer would be unable to recover even his variable costs; no matter what quantity he might choose to produce, he would be continually losing all his fixed costs and something more. Even so a producer may sometimes continue to operate — temporarily. Hoping for better times in the future, the producer may choose to bear a present loss in order to keep his plant from deteriorating or to hold together his organization and an experienced labor force. Ordinarily an enterprise that is losing more than its fixed charges will very soon find it best to close down.

Short run and long run. The distinction between the costs which a producer must cover in the short run and those which he must cover in the long run is of great importance for many kinds of economic analysis. For instance in the course of a depression most producers will operate for considerable periods of time at "less than cost." Nevertheless they do continue to operate, which may be taken as a sign that their variable costs at least are covered, or are very nearly covered, by their current operations. The distinction

between the short run and the long run is especially important for a modern industrial community, where so many businesses have enormously heavy fixed costs. For instance in the great depression following 1929 the railroads were very hard hit. Their business dropped off disastrously. Variable or operating costs were comparatively rigid, and many large systems were forced into bankruptcy; yet there were practically no "class I railroads" which actually incurred operating losses; and of those that did incur operating losses it is probable that there were none for which variable costs at least were not covered. A railroad's operating costs include such items as maintenance, depreciation, and administrative overhead. These expenses would have continued even if the railroads had shut down entirely, and they cannot therefore be classified as variable costs. We see then that even in the case of this "sick" industry, in the trough of the great depression, it would not have been profitable to have ceased operations entirely, because the railroads' losses would in that case have been even greater than they were.

Conclusion. We have now investigated the various cost schedules and have considered their relationship to price under conditions of pure competition. From this study three propositions concerning output and price have emerged. (1) So long as the firm continues in operation under pure competition it will seek to produce that output for which marginal cost is just equal to price. This output will maximize its profits when price exceeds average total cost and will minimize its losses when price is below average total cost. (2) A firm will not continue production in the long run unless it receives a price that is sufficient to cover average total costs. (3) A firm will continue operations in the short run so long as the price exceeds average variable costs, and it may continue to operate for a short time even if price falls slightly below average variable costs. These three propositions describe the way in which the firm's cost schedules will determine the amount that it will produce, always assuming that it can sell any amount which it chooses at the going market price.

Under imperfect competition: The seller and the demand for his product. We must now examine the position of a typical firm

which is engaged in imperfect competition. Here the elasticity of demand for the product of a single seller is not perfect.¹

As we have observed, this is the situation of most industrial producers in the United States; either the competing products have certain individualities which appeal to different customers, or the number of competitors is so small that the individual's selling policy may have an appreciable effect on the prices and the sales of other competitors. The demand for his product does not present itself to him in the form of a going market price at which he must sell or refuse to sell. Instead he faces some sort of demand schedule, which is a way of saying that the number of units he will sell depends upon the price which he sets. He has a price problem.

The producer who is selling in an imperfect market has a position of privilege, but it relates to the seller's side of the market only. Since he usually sells in what we have defined as a free market, he may fix either the price at which he will sell or the quantity, but not both. In case he takes the first course, demand will determine the maximum quantity he can sell. In the other case, demand will determine the maximum price he can charge. In any case his problem is that of discovering the most profitable price or quantity to set in view of the shape of the demand curve for his individual product and of his various cost schedules, such as we studied in the last chapter. These principles and the considerations which guide the producer in making his decisions as to what price to charge or what quantity to sell may be made clear by use of the graphical aids with which the reader is now familiar.

Getting the maximum profit. Some people think that under imperfect competition a producer will charge the highest possible price and will try to sell the largest possible quantity to the greatest number of buyers in order to make the largest possible profit. A little consideration will show that he cannot do all these things at the same time. If he charges the highest possible price, he will transact in most cases only a very small amount of business. In order to transact the largest possible amount of business he must in most cases be prepared to accept a very low price. In neither event

¹ See Chapter IX for definitions and descriptions of pure and imperfect competition.

is he likely to make the largest possible profit. To get the most for his product he generally has to adopt some sort of golden mean; *i.e.*, charge neither too much nor too little.

Suppose the demand for the product of the firm is as represented by the curve dd' in Figure 16. cc' is the representation of the schedule of average total costs. If this producer charges a very high price, such as pa , his sales will be only Oa . If he hopes to sell the very large amount Oa' , he must be prepared to accept the very low

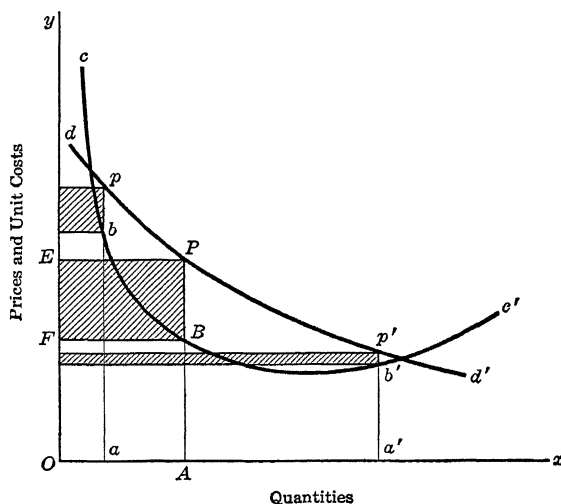


FIG. 16

price $p'a'$. In the first case, his total costs are represented by the rectangle with its base on the horizontal axis and its upper right-hand corner at point b on the average cost curve, for the area of this rectangle is equal to Oa (total quantity) times ba (average cost). Under the same circumstances his gross receipts are represented by the rectangle with right-hand corners at points a and p , this area being equal to Oa (total quantity) times pa (price). His net profits then are represented by the comparatively small shaded rectangle with corners at p and b , for this area is equal to the difference between the gross receipts rectangle and the total cost rectangle.

Sizing up the areas of rectangles with upper right-hand corners at b' and p' respectively, we see that selling the quantity Oa' at price

$p'a'$ would again likewise result in comparatively small profit, as represented by the narrow shaded area with right-hand corners at p' and b' .

But now let us consider an intermediate price such as PA . Here the profit rectangle is much greater than in either of the previous cases. If the profit of the firm is to be made a maximum, the problem, translated into graphical terms, is to select a price such that the area of the profit rectangle, $EPBF$, will be as great as possible. In numerical terms, this means finding a price such as \$1.00 in the following schedule :

<i>Price</i>	<i>Sales</i>	<i>Gross receipts</i>	<i>Total cost</i>	<i>Net profit</i>
\$1.20	6,000	\$ 7,200	\$ 3,000	\$4,200
1.10	8,000	8,800	4,000	4,800
1.00	10,000	10,000	5,000	5,000
0.90	12,000	10,800	6,000	4,800
0.80	14,000	11,200	7,000	4,200
0.70	16,000	11,200	8,000	3,200
0.60	18,000	10,800	9,000	1,800
0.50	20,000	10,000	10,000	0

The question now arises : Is there not some general rule or principle which will describe the attainment of maximum profit ? It must be recognized that the producer needs to know the demand for his product. Sometimes his knowledge is very vague — a matter of guesswork. But assuming that he has some idea of the demand for his product, or that he can find out through experiment, how can we describe the most profitable price adjustment that he can make ? An answer is found in a very simple relationship between *marginal revenue* and *marginal cost*. The concept of *marginal cost* has already been discussed in Chapter IX ; *marginal revenue* is a complementary idea.

Marginal revenue. Suppose 90,000 pairs of a certain make of shoes can be sold at a price of \$4.50, but to sell 100,000 pairs the price has to be reduced to \$4.25. The total revenue or gross receipts from the sale of these respective amounts are as follows :

100,000 pairs @ \$4.25	\$425,000
90,000 pairs @ \$4.50	405,000
Difference in revenue	\$ 20,000

Rate of difference in revenue :
 $\$20,000 \div 10,000 = \2.00

In this example \$2.00 is said to be the marginal revenue derived from the sale of 10,000 pairs of shoes, in accordance with the following definition: *Marginal revenue is the amount per unit by which, under given demand conditions, a small change in volume of sales changes total revenue.* The phrase "a small change" implies the smallest change in sales for which, as a practical matter, a corresponding change in price can be calculated or estimated. As our mathematical illustration clearly shows, the marginal revenue is calculated by dividing this "small change in sales" (10,000 units) into the corresponding increase in revenue (\$20,000). Marginal revenue is therefore strictly a *rate*; and it is sometimes defined as "the rate at which, under given demand conditions, a small change in volume of sales changes total revenue." Obviously this definition and the one we have chosen are identical in meaning.

The pure mathematical concept of marginal revenue is based upon the assumption of infinitesimal changes in sales, price, and revenue. For purposes of exposition it may sometimes be convenient to consider theoretical changes of one unit. Marginal revenue is then thought of as the amount added to total revenue by the sale of one additional article. Of course it must always be remembered that in practice this result can seldom if ever be measured. No manufacturer of shoes could ever estimate the change in total revenue that would occur if sales were increased from 99,999 pairs to 100,000 pairs of shoes.

To complete the concept of marginal revenue, we must add that if the small change in sales takes place through addition, the change per unit in revenue is said to be the marginal revenue of the quantity which includes the added units.

If a small change in sales can be effected without a change in price, and without change in the general conditions of demand, price and marginal revenue are the same. For example, if 100,000 pairs of shoes can be sold at the same price as 90,000 pairs, we may have:

100,000 pairs @ \$4.50	\$450,000
90,000 pairs @ \$4.50	405,000
Difference in revenue	\$ 45,000

Rate of difference in revenue:

$$\$45,000 \div 10,000 = \$4.50$$

The price is also \$4.50

Since there is no change in price, the sale of each additional pair brings in an additional \$4.50, and the rate of change in revenue is \$4.50 throughout. This is obviously the situation for a producer selling in a perfect market; his marginal revenue is always equal to the price, since any change in the individual's offerings has no appreciable effect upon the price.

In the case of a producer selling in an imperfectly competitive market however a change in sales has to be accompanied by a change in price. Thus the change in total revenue is made up of two component parts; it will be increased by the revenue from the sale of the additional units, but it will be decreased as a result of the reduction in price of the number of units which was previously being sold.

Marginal revenue, when not equal to the price, is always less than the price, never greater. The reason is that under these conditions, an increase in sales has to be accompanied by a decrease in price under the same demand conditions. When total revenue varies inversely with the quantity sold, marginal revenue is considered negative. Suppose, for example, that 90,000 pairs of shoes can be sold at \$4.50, but that to sell 100,000 pairs the price has to be reduced to \$4.00. We have then:

100,000 pairs @ \$4.00	\$400,000
90,000 pairs @ \$4.50	405,000
Difference in revenue	<u>\$-5,000</u>

Rate of difference in revenue:

$$\$-5,000 \div 10,000 = \$-0.50$$

A producer will of course never, if he can help it, expand his sales up to the point at which marginal revenue is negative, since if he found himself in this situation, it would mean that his gross receipts could be increased by decreasing the volume of his sales. This is by no means an impossible or even unusual position. Indeed whenever the elasticity of demand for the individual producer's product is less than unity, he will be able to expand sales only at the expense of such large price reductions that his total receipts will be reduced.

Marginal revenue and marginal cost: In general. Price above average cost. Whenever a producer discovers that his marginal revenue is greater than his marginal cost, it is clear that he will gain

by at least a slight increase in production, because such an increase will add more to his revenue than to his cost. As a result of this adjustment, marginal revenue itself will decline and marginal cost may rise so that a further adjustment will presumably add less to his profits and might even reduce them. But small adjustments which involve a cut in price and consequent increase in sales will always offer the possibility of additional profits so long as marginal revenue exceeds marginal cost. In graphical terms there is the possibility of increasing the area of the profit rectangle (see Figure 16).

When marginal revenue is equal to marginal cost, so that any further increase in production would increase cost as much as it would increase receipts, additional sales will no longer increase the producer's profit. By the same reasoning, if marginal revenue is less than marginal cost, an increase in production will decrease profits, and total profits can be increased only if production is reduced. In this situation it is to the producer's interest to raise his price (if he sells in a market where price is set by the seller) or directly to limit production. Thus we reach the conclusion that *profits are at a maximum when marginal revenue and marginal cost are equal*.

For the sake of a simple numerical illustration let us assume that marginal revenue is based upon changes of 2,000 units in sales and that marginal cost is always \$0.50, as in the schedule below.

Suppose the producer considers selling 6,000 units at price \$1.20. It pays him to increase the amount to 8,000 units even though the price has to be reduced to \$1.10. His gross receipts with this reduction in price increase \$1,600; *i.e.*, at the rate of \$0.80 for each addi-

Price	Sales	Gross receipts	Marginal revenue	Marginal cost
\$1.20	6,000	\$ 7,200	—	—
1.10	8,000	8,800	\$0.80	\$0.50
1.00	10,000	10,000	0.60	0.50
0.90	12,000	10,800	0.40	0.50
0.80	14,000	11,200	0.20	0.50
0.70	16,000	11,200	0.00	0.50
0.60	18,000	10,800	-0.20	0.50
0.50	20,000	10,000	-0.40	0.50

tional unit. This more than covers the marginal cost of \$0.50. For similar reasons it pays him to increase production from 8,000 to 10,000 units. In this instance marginal revenue is \$0.60 as com-

pared with marginal cost of \$0.50. It does not pay him however to reduce the price to \$0.90 and sell 12,000 units. Here marginal revenue is only \$0.40.

If price were reduced to this figure, the final units sold would be adding \$0.10 less to revenue than to cost. The producer would still be making a profit to be sure, but not so much as he might. On the other hand, selling 10,000 units at price \$1.00, he has a marginal revenue a little greater than marginal cost. If he were disposed to alter his sales by smaller jumps, he would probably find the ideal output around 11,000 units. With finer calculations of marginal

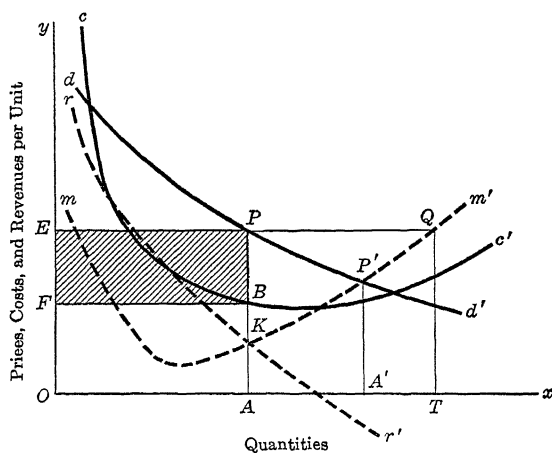


FIG. 17

items, he would be able to find an exact equality between marginal revenues and marginal costs somewhere in this neighborhood. In general analysis of price under imperfect competition it is customary to assume such exact possibilities, not that they are realized in practice, but because minor discrepancies are not important from the general point of view.

Such is the assumption in our analysis of Figure 17 above, which represents more general conditions of demand and cost of production. cc' is the average cost curve for a product sold in an imperfect market. dd' is the demand curve. rr' is the representation of a schedule of marginal revenues. dd' and rr' correspond; both relate to the same conditions of demand. The marginal rev-

enue schedule is derived from the demand schedule. mm' is the firm's marginal cost curve. It intersects the curve rr' at point K .¹ The perpendicular distance AK represents the marginal revenue derived from the sale of the quantity OA and also the marginal cost of this same quantity OA . Here marginal revenue and marginal cost are equal. It pays the firm to expand its operations up to this point. The additional revenue from any small final increase in production covers the additional cost. Furthermore the additional costs of other increases which may have been planned or considered up to this point are more than covered by the additional revenue, due allowance being made for necessary reductions in price, as shown in the diagram. It does not pay the producer however to produce and sell more than OA , for, as shown by the two curves, additional cost in this case will not be covered by additional revenue. If measures of marginal cost and marginal revenue are represented in the diagram with great accuracy, being based on very minute changes, then OA is the exact quantity which will give the firm maximum profits. Otherwise this is the approximate quantity.

With the conditions of demand as represented by the curve dd' , it is clear that the quantity OA can be sold at the price PA determined by extending the perpendicular AK up to the demand curve at P . PA is the price the producer will charge and the price at which his profits will be at the maximum. Taking into account the cost curve cc' , we have the same graphical representation of gross receipts, total costs, and maximum net profits as has already been portrayed in Figure 16. That is, the firm's profit resulting from the price PA is measured by the shaded rectangle $EFBP$.

Price below average cost. In the examples we have so far discussed under imperfect competition, we have assumed that it will be possible for the firm to earn a profit over and above its total costs for a very considerable range of outputs. Graphically this assumption was implied in Figure 17 by the fact that the demand

¹ If the marginal revenue curve intersects the *descending* branch of the marginal cost curve instead of its ascending branch as in Figure 17, the intersection still indicates the most profitable volume of operations for the monopoly. The only proviso is that the marginal revenue curve must be higher than the marginal cost curve to the left of their intersection. Otherwise the intersection would indicate the least profitable volume of operations, not the most profitable.

curve dd' was above the average cost curve cc' throughout a considerable part of its length. Although the firm must produce the output OA in order to maximize its profits, it will be able to earn some profit if it produces any output for which price exceeds average cost. In order therefore to round out our consideration of the behaviour of producers under imperfect competition, we must take account of the possibility that they will not always be so favorably situated. Suppose for instance that the demand curve dd' did not

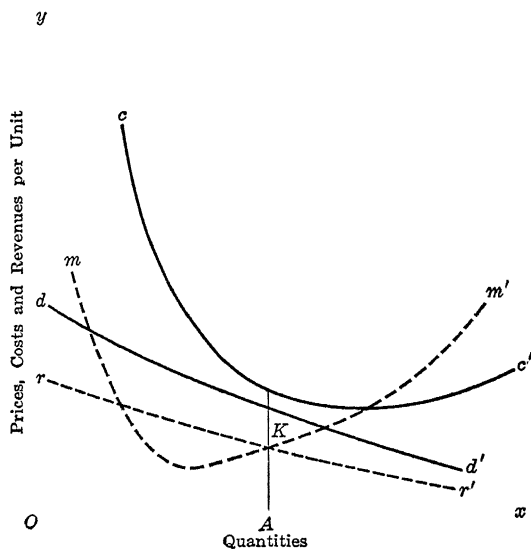


FIG. 18

intersect the average cost curve cc' at any point but lay below the latter entirely. (See Figure 18.) This would imply that the firm might lose more money for some outputs than for others, but that it could not under any circumstances cover its costs. What will be the significance in this case of the point of intersection of the marginal revenue and marginal cost curves, and what action is our firm likely to take?

The first of these questions may be easily answered if we apply exactly the same reasoning that has been developed in the preceding paragraphs. Although it is obvious that the firm's profits cannot be maximized by the equating of marginal cost and marginal

revenue, nevertheless the firm's losses will be minimized when the firm is producing this output. A moment's consideration makes it clear why this is bound to be the case. If the firm reduces its output below the point at which marginal revenue and marginal cost are equal then it will reduce its revenues by more than its total costs; consequently, its losses will be increased. If it expands output into the region where marginal cost exceeds marginal revenue then more will be added to costs than to receipts and again losses will be increased. We may safely conclude then that *so long as the firm remains in business it will tend to produce the output at which marginal cost is equal to marginal revenue so as to minimize its losses.*

This however does not provide a complete answer to the second of the above questions. It will readily occur to the reader that a firm cannot continue to operate at a loss indefinitely under conditions of pure competition. But by the same token, a firm operating under imperfect competition may remain in business for a considerable period of time operating at a loss if variable costs are covered, that is to say, if it is earning something, however small, toward meeting its fixed costs. This reasoning here applies in exactly the same fashion as to a firm that is operating under conditions of pure competition.

Conclusion. We may now sum up the results of our investigation of the relationship between cost and the price and output policies of a firm in three propositions, which are analogous to those which emerged from our study of pure competition. (1) So long as a firm continues in operation under imperfect competition it will tend to produce that output by which marginal cost is just equal to marginal revenue. (2) The firm will not continue production in the long run unless it receives a price that is sufficient to cover average total costs. (3) It will continue operations in the short run so long as price exceeds average variable costs.

Imperfect competition and the elasticity of demand. In our discussion of the relationship between cost, price, and output for a firm selling under conditions of pure competition, we concluded that the firm would always tend to push its sales to the point at which marginal cost was equal to price, the reason being that, as the firm

could sell any amount that it wished without appreciably affecting price, it would maximize its profits by seeking this adjustment. For a firm selling in an imperfectly competitive market the general rule at which we have arrived is that it will tend to push its sales until marginal revenue is equal to marginal cost. This suggests that if we are to make any generalizations about the effects of imperfect competition as compared with pure competition, we must try to discover what it is that determines the difference between marginal revenue and price. The answer is not far to seek. Under imperfect competition marginal revenue is, we have seen, always lower than price, because a firm's total revenue is increased when output expands by the receipts from the added units sold, but this increase in revenue is offset by the decrease in price which is necessary in order to secure the increase in sales and which decreases the revenue from the goods that could have been sold without a lowering of price. It is clear that, if price must be reduced very largely in order to secure an increase in sales, marginal revenue will fall far short of price; whereas if a very small reduction in price will secure a large increase in sales, the marginal revenue will be nearly equal to price. In fact we have noted that under conditions of pure competition, where sales may be expanded without any decline in price, marginal revenue is equal to price.

This suggests that under conditions of imperfect competition it is the elasticity of demand that is crucial. If demand is highly inelastic, it will require a large reduction in price to secure a given increase in sales, and thus marginal revenue, for any given output, will be much less than price. Therefore if the elasticity of demand for the product of an individual firm is extremely low, price under conditions of imperfect competition will far exceed the marginal cost which will be equated to marginal revenue. If on the other hand, demand is highly elastic, then the difference between pure and imperfect competition begins to disappear, because marginal revenue will be only slightly less than price, and thus in equilibrium marginal costs (which will be equated to marginal revenue) will be very little less than price. We may sum up the discussion of the relationship between price and cost under conditions of imperfect competition by saying that the price set by a firm will depend upon

marginal costs and upon the elasticity of demand for its individual output. If demand is elastic, price will be close to marginal cost; if demand is inelastic, price will be well above marginal cost; in any case, these will be the two significant elements in the situation.

Imperfect competition and monopoly. The foregoing discussion of imperfect competition leads naturally to the subject of monopoly. There are monopolies controlling the sources of certain products of nature, such as the Carlsbad mineral waters, the diamonds of Africa, and the anthracite coal of Pennsylvania. The possessor of any unique non-reproducible article, such as a famous work of art, is of course in a monopoly position. A famous actor or musician has a monopoly in the personal services which the public demands of him alone. The United States government has a monopoly of the service of carrying letters. Patents and copyrights give monopolies to inventors and authors, and city governments grant monopolies to corporations in the form of exclusive franchises to operate street railways, to furnish water, gas, and electricity, etc. Finally there are those great corporations and combinations, in fields in which the advantages of large-scale production and combination are especially strong, which have come to positions at least approaching monopoly in their particular fields.

Our investigation of imperfect competition readily suggests that actually the monopolist differs in degree rather than in kind from the ordinary industrialist. The local electric light company may be a monopolist because no one else may sell electric current in the same district, but there is a demand curve for the product of the local utility just as there is a demand curve for the output of Mr. Ford's company. Furthermore recent experience would appear to indicate that the demand for electric current is highly elastic. Therefore the president of the local utility has to make a decision of precisely the same kind as that which Mr. Ford must make. The most difficult question which each of these men must answer concerns the increase in sales which will result from any given reduction in the price of what he sells. Each must decide what price to set or what output to throw on the market with reference to exactly the same criteria.

The only real distinction between a monopolist and a business man selling in an imperfectly competitive market is that the monopolist sells a product which has no close and exact substitute, whereas the ordinary competitive business man, although he may sell a highly individualized product produced by no one else, nevertheless must as a rule meet the competition of close substitutes. For instance although there is only one manufacturer of the Ford car — Henry Ford himself — there are other cars which are nearly its equivalent in quality and in price. The fact is that these substitutes also limit Mr. Ford's power with regard to the setting of price. On the other hand, the local water company does not have to compete with any close equivalent and may thus properly be called a monopoly.

From this distinction we may infer that the demand curve for the monopolist's product is usually less elastic than that facing the business man who has competitors but sells in an imperfectly competitive market. If the power company increases its rates, its consumers generally have no alternative but to economize. If Henry Ford raises his prices, his customers may turn to Plymouth and Chevrolet. Though ultimately still only a difference in degree, this is the significant distinction between the two for the purposes of economic analysis.

Monopoly and the public interest. The seriousness of a monopoly situation is mostly dependent upon whether the monopoly controlled commodity is a luxury or a necessity. We have already had occasion to note that the demand for necessities is usually very inelastic. If therefore the necessity is produced or sold by a monopoly, it will pay the monopolist to set his price extremely high, because marginal revenue will be much lower than price (in fact it will probably be negative for most outputs) and thus when the monopolist has equated marginal cost and marginal revenue price will be high. On the other hand, monopoly control of a luxury is no great menace. The monopoly will, in its own interest, generally refrain from charging an exorbitant price. And if its price is excessive, the people can do without this particular article.

Anticipating our later study of the subject of monopoly, it may be stated here that in general the public interest is better served

when production is carried on under the régime of competition. There are certain exceptions, some self-evident, others not so obvious. Personal monopolies, such as those enjoyed by the popular singer or actor, are inevitable and can do the public little harm. The monopoly granted by the patent or copyright is generally in the public interest. And there are certain lines of industry, such as railroading, telephone, telegraph, the furnishing of water, gas, electricity, etc., which from their economic nature appear destined to be monopolies whether or no. In such cases it is futile to put our trust in competition which cannot endure; monopoly must be accepted and made the best of. This topic will present itself for our further study in a later chapter.

Public opinion today generally holds that abuse of monopoly power, particularly in the necessities of life, should so far as possible be prevented in order to safeguard the public interest and that when monopoly is inevitable government regulation should be relied upon to ensure to the public good service and reasonable prices. For this general opinion there is firm foundation in economic principles.

EXERCISES

1. Assume the following conditions of cost for the production of automobile heaters in the Acme Products plant :

<i>Output</i>	<i>Fixed cost</i>	<i>Variable cost</i>	<i>Total cost</i>	<i>Average variable cost</i>	<i>Average total cost</i>	<i>Marginal cost</i>
10,000	\$150,000	\$ 75,000	\$225,000	\$7.50	\$22.50	\$ 7.50
20,000	150,000	145,000	295,000	7.25	14.75	7.00
30,000	150,000	216,000	366,000	7.20	12.20	7.10
40,000	150,000	288,500	438,500	7.21	10.96	7.25
50,000	150,000	363,500	513,500	7.27	10.27	7.50
60,000	150,000	442,500	592,500	7.38	9.88	7.90
70,000	150,000	527,500	677,500	7.54	9.68	8.50
80,000	150,000	619,000	769,000	7.74	9.61	9.15
90,000	150,000	719,000	869,000	7.99	9.66	10.00
100,000	150,000	832,000	982,000	8.32	9.82	11.30

- Plot on a sheet of graph paper the curves of average variable cost, average total cost, and marginal cost.
- Approximately what is the optimum size, or point of maximum efficiency, for this concern?

- (c) Assume that the product is sold under conditions of pure competition. If the best market price obtainable for the output were \$7.10 per unit, would you expect the concern to produce any heaters? If so, how many? Explain.
- (d) If the best market price obtainable for the output were \$8.50 per unit, would you expect the concern to produce any heaters? If so, how many? What would be the loss per unit, and the total loss, sustained? What would be the loss if the plant closed entirely?
- (e) If a price of \$10.00 per unit could be obtained for the output, how many heaters would be produced? What would be the profit per unit, and the total profit, under such conditions? Draw the rectangles which correspond to total cost, total receipts, and total profit, in this case.

2. Now assume that the product of this concern is not homogeneous with that of its competitors, but is distinguished through branding or special features of construction. The table below represents the demand schedule for Acme heaters, sold under conditions of imperfect competition :

<i>Prices</i>	<i>Quantities</i>
\$15.00	10,000
14.00	20,000
13.00	30,000
12.00	40,000
11.00	50,000
10.00	60,000
9.00	70,000
8.00	80,000
7.00	90,000
6.00	100,000

- (a) For each price and amount of sales in the demand schedule, compute the total receipts that would be obtained from selling that amount at the corresponding price.
 - (b) By reference to the figures of total cost in Exercise 1, compute the net profit or loss at each possible price and output.
 - (c) At what price and output are the profits of the concern maximized?
3. (a) From the demand schedule in Exercise 2, compute the figures for marginal revenue.
- (b) Plot on a sheet of graph paper the demand curve, and the marginal revenue curve.
 - (c) From the cost schedule in Exercise 1, plot again the curves of average total cost and marginal cost (four curves on the same graph).
 - (d) What is the best output for the concern to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?
 - (e) What is the profit per unit and the total profit under these conditions? Draw the rectangle which corresponds to total profit.

4. Assume now a different demand schedule for Acme heaters, as indicated below :

<i>Prices</i>	<i>Quantities</i>
\$11.50	10,000
11.00	20,000
10.50	30,000
10.00	40,000
9.50	50,000
9.00	60,000
8.50	70,000
8.00	80,000
7.50	90,000
7.00	100,000

- (a) For each price and amount of sales in the demand schedule compute the total receipts, and the marginal revenue, that would be obtained.
- (b) Plot on a sheet of graph paper the demand curve, and the marginal revenue curve.
- (c) From the cost schedule in Exercise 1, plot again the curves of average total cost and marginal cost (four curves on the same graph).
- (d) What is the best output for the concern to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?
- (e) What is the loss per unit, and the total loss under these conditions? Draw the rectangle which corresponds to total loss. What would be the loss if the plant closed entirely?

XII

SUPPLY

In the last two chapters we have studied the conditions which influence individual sellers in the determination of their selling policies. We have seen that foremost among these conditions is cost of production. We must now carry our investigation a step further toward discovering the manner in which prices in any trade or industry are determined by the actions of the sellers together with the demand of the buyers. In the present chapter and the next we shall be concerned with a study of the seller's side of the market under conditions of pure competition.

Individual supply: Normal condition. For the study of the seller's actions under such conditions the concept of supply is fundamental. *An individual's supply of any good is a schedule of the respective quantities of that good which he is ready to offer for sale at all possible prices.* The reader will observe that this definition corresponds precisely to the definition of individual demand, which was developed in Chapter VIII. The ordinary relation between price and the quantity offered for sale by an individual follows readily from the principles which were developed in the last two chapters. As long as any industrial enterprise continues to operate under conditions of pure competition, the major portion of its rising marginal cost curve (see, for example, the segment lm' in Figure 13, page 226) becomes the basis of its supply curve. If the management, in its possible decisions regarding offerings and output, accurately and consistently follows the information pictured in the rising branch of a marginal cost curve, its supply curve and the indicated portion of the marginal cost curve exactly coincide. Though the two curves exactly coincide however, it is still necessary to preserve a certain distinction between them. The marginal cost curve shows a relationship between marginal cost and the number of units to be produced. In this case the marginal cost depends upon the quantity of goods. The quantity of goods is, in

the language of mathematics, the independent variable. The supply curve of an individual firm, though based on a marginal cost curve, shows a relationship between price and the number of units to be offered for sale. In this case the quantity of goods depends upon the price; the quantity of goods is not the independent but the dependent variable.

We have learned that when competition is pure each firm is free to produce whatever quantity it wishes. At any given time each firm will therefore be producing the quantity that will maximize its profits or at least minimize its losses, which means that production will always be pushed beyond the point where the marginal cost curve begins to slope upward. In other words, the firm cannot possibly be in equilibrium unless its marginal cost curve is sloping upward. These rising marginal cost curves are, as we have seen, the supply curves of the several individual firms. They illustrate the general principle that, in a perfect market, the quantity of a good offered by an individual seller normally varies directly with the price.

An exception. Lest we overlook any conceivable case, notice should be taken of the possibility of another form of supply curve, which may occur in connection with the offer of goods or services by extremely improvident sellers. Thus traders dealing with savages have sometimes sought to stimulate the natives to increased activity in bringing in their products by the offer of higher prices, only to find the result disconcertingly contrary to the normal law of supply. The native was content when he had earned a certain amount; thereafter he proposed to knock off and enjoy his leisure. The higher the price, the fewer the goods he had to sell to reach his goal. His supply curve, after ascending normally for a while, turned back on itself (see Figure 19). As we observed when studying similar cases in connection with the law of demand, these special cases do not affect the validity of the general relation between price and the quantity offered by an individual seller.

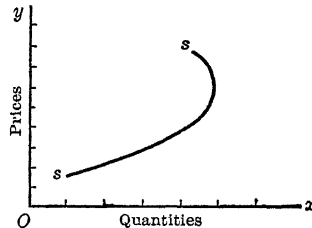


FIG. 19

Total supply: Definition. Any individual supply curve (or schedule) governed by and related to marginal cost combines with other individual supply curves (or schedules) to form the total market supply. The process is similar to that already described in combining individual demands into the total market demand.¹

The supply of any good is a schedule of the respective quantities of that good which people are ready to offer for sale at all possible prices. This definition evidently corresponds precisely to the definition of demand. So close indeed is the analogy between demand and supply that a few words only at this point will suffice to set forth the formal character of the latter. The phrase "ready to offer for sale" implies both ability and willingness. Supply, like demand, relates to a particular place and a particular time; in strictest theory to an instant of time, though for most purposes the time is an appreciable period — a few minutes, an hour, a day, a week, etc., during which the respective quantities would presumably be offered if various assumed prices should prevail. A supply schedule, like a demand schedule, indicates a series of simultaneous possibilities — on the other side of the market however.

Supply and stock. Much discussion of the various problems of price goes astray through confusing supply with the stock of goods on hand. For example what is popularly known as the "visible supply" of various grains in the United States is reported monthly by the United States Census Bureau. On page 253 are the figures for wheat and corn during the first eleven months of 1938. These are important quantities, and they have a real bearing upon supply, but they are not the supply of wheat or corn in any market.

The stock of wheat of various grades in the elevators in the neighborhood of Chicago is not the supply of wheat on the Chicago Board of Trade. The supply is the schedule of the respective numbers of bushels of wheat that people are not only able but willing to offer for sale at various prices at a particular time. These quantities will not be the same as the quantity of wheat in the Chicago elevators. In fact, while there is a relation between the quantity of goods on hand and the supply, this relation will often appear not to be very close. The quantity of wheat offered for

¹ See page 160, Chapter VIII.

sale at Chicago at some particular price may be far less than the wheat on hand in the elevators. Much of that wheat the holders may not be willing to sell at that price. On the other hand, the quantity offered might be more than the total stock in the elevators.

VISIBLE SUPPLY OF WHEAT AND CORN¹

1938	Wheat		Corn
	UNITED STATES	CANADA	UNITED STATES
January	79,203,000	50,088,000	41,092,000
February	66,467,000	45,528,000	39,000,000
March	54,426,000	43,379,000	43,227,000
April	43,191,000	41,029,000	40,704,000
May	33,816,000	31,690,000	25,916,000
June	28,333,000	25,043,000	23,674,000
July	96,389,000	18,726,000	15,004,000
August	133,725,000	52,429,000	10,489,000
September	139,273,000	150,635,000	9,899,000
October	141,914,000	173,542,000	23,081,000
November	136,204,000	162,379,000	46,645,000

The Chicago market may be drawing upon wheat stored in other places. In fact the quantity offered might conceivably be greater than the total stock in existence. This surprising state of affairs occurs now and then in speculative markets. On May 9, 1901, on the New York Stock Exchange, it was suddenly discovered that sellers had contracted to deliver more shares of the stock of the Northern Pacific Railway Company than could possibly be bought or borrowed by them. A disastrous crisis was the result. Of course the people were offering shares which they did not then possess; they hoped to acquire them in time to deliver. But that does not alter the fact that so many shares were actually offered. In 1917, during the World War, there occurred a "corner" in wheat on the Chicago Board of Trade. At one time the price was \$3.25 a bushel, and it was stated by a well-known authority that "we are getting into a serious situation. Everybody is long, and there is not available the wheat to deliver to those who have bought it."²

¹ *Survey of Current Business*, January, 1939, pp. 42 and 43.

² J. A. Patten, "In the Wheat Pit," *Saturday Evening Post*, Nov. 5, 1927, p. 177.

Supply pertains to pure competition. Before proceeding to employ the concept of supply in the analysis of the seller's side of the market, it will be wise to emphasize the fact that the concept of supply, together with supply schedules and supply curves, as we have defined it, analogous to the corresponding concept of demand, has no meaning except when the amount of a commodity that will be offered in the market depends exclusively upon the market price; *i.e.*, the quantity offered is a function of price alone. This is a condition which is to be found only when competition is pure.

In any imperfect market the amount that a given firm will decide to produce and sell will not depend simply on the price at which it can dispose of its output. It will depend also upon such other factors as the selling costs which the firm must incur in order to sell a given output at a given price and upon the possibility of selling a smaller output at a higher price or a larger output at a slightly lower price. In other words, under a régime of imperfect competition the amount that a firm will "offer" depends upon selling costs and the elasticity of demand as well as upon the price at which the output will be disposed of. Clearly under these conditions it would be quite impossible to draw up a supply schedule showing the various amounts that will be offered for sale at different prices. A certain price may stimulate a very large output if firms can sell all at that price. But the possibility of selling a certain output at that price will not induce individual firms to establish the price and sell the output which the market will absorb if it is also possible for them to sell practically as large an output at a much higher price or to sell a much larger output at a slightly lower price. In fact as soon as we get a little deeper into this question, we can see that it is impossible to speak without qualification of a certain "market price" as the sole force inducing producers to "offer" a certain imminent output if they are selling in an imperfectly competitive market. For in the typical market of this sort the producers themselves set the prices of their products, and in so doing they will have to take account of the whole demand curves for their own products rather than an established market price. Thus in such a market the notion of a supply curve has no meaning.

If this point is fully grasped the reader will avoid much confu-

sion. It is extremely useful to sum up a great deal of information about, let us say, the wheat industry in the form of a supply curve for wheat; indeed it is impossible to explain the determination of the price of wheat except through the use of this device. But it is impossible, in the strict sense in which we define the term, to talk about the supply curve of cheap automobiles, and if we try to do so we shall derive a false picture of the forces which determine price in an imperfectly competitive market.

Significance of the concept of supply. It does not follow that the concept of supply is of only slight value. Besides the many important lines of production, such as wheat and cotton, where virtually pure competition prevails and the concept of supply is essential to an understanding of price, there are many other industries, where, although competition may not be quite so pure, that condition is still approached. Many such industries and many markets include a fairly large number of competitors. In a given place and time it may seem as if competition is limited to very few, but indirectly there may be competition among many. Suppose a number of stores, A, B, C, D, E, . . . N, each operating under independent management. A seems to compete only with B and C; but C also competes with D and E. A then is indirectly competing with D and E. So on down the line; all stores from A to N may be directly or indirectly competing with each other. The concept of supply as a schedule of schedules offers unlimited opportunity for further investigation. In the meantime the simpler sort of schedule affords an insight into the working of the forces of demand and supply which otherwise we might not get at all. Incidentally there are cases in which sellers ignore the possibility of changes in other prices and concentrate for the time being on the price of the good they have to sell. There are also cases in which, temporarily at least, there really is no significant change in other prices, or no significant change in the relations between prices. Here a simple supply schedule is of immediate significance, provided there is at least a fairly large number of sellers of about the same size and importance.

Relation of quantity to price: Long and short run. As we have seen, the normal relation between price and quantity offered by an

individual seller is that the quantity varies directly with the price. It might appear that the combination of the several individual supplies would lead to a similar relationship for the whole market. This is indeed the result when we are considering the "short run" only — what we have called, in Chapter IX, the "immediate sales problem." If we are thinking only of the quantities that sellers could be induced to part with at once from their existing stocks, it is true in practically all cases that the higher the price — the greater will be the quantity offered by sellers.

But when we turn our attention to the "long run" or "remote sales problem," the matter is not quite so simple. Here we are interested in the quantities that will be produced for sale over a period of time, involving consideration by the sellers of their future costs of production. The shape of the long run total supply curve of any industry will be affected by the cost conditions that prevail in that industry.

Increasing costs. Let us start our inquiry into the law of supply by supposing that at a given time production had all to come from the establishments then in operation. It is clear that under this assumption an increased product could be obtained only by offering a higher price, permitting each of the individual firms to increase its output at the expense of higher marginal cost, which for all producers would be equated with the price. We should then have an upward sloping total supply curve, indicating that the higher the price the greater would be the number of units offered for sale.

Let us next consider another condition, much more common in actual business, which would also give rise to the upward sloping total supply curve. Here production is not limited to the existing firms, but if new firms enter their average costs, when they are operating at the most efficient size, will be higher than the existing price. In other words, the existing price will be below the minimum points on their average cost curves. That is the reason they are not already in the market. But if prices were high enough they would enter and offer for sale such quantities of the product as would equate their marginal costs with the price. Of course each of the firms already in the market would increase its production till its

marginal cost and revenue were similarly equated. We then have a situation with an upward sloping supply curve, the number of competitors and the amount produced by each varying directly with the price.

The supply situation just described, whether it arises from the lack of new firms able to enter the industry or from the fact that new firms can enter only at higher costs than those of the existing firms, may be described as a condition of increasing costs in the industry.

The condition which is generally responsible for the state of increasing costs is scarcity of one of the factors of production which all firms in the industry require. If for instance a particular kind of skilled labor needed, let us say, in the electrical industry is extremely scarce, then it will be impossible for a new firm to enter the industry with costs substantially lower than those of existing firms. If such a new firm were to enter the industry, its attempts to secure control of a supply of scarce skilled labor would presumably make such labor more expensive to all producers and would force all producers in the industry to resort to labor of inferior quality. Thus the appearance of the new firm would be marked by an upward shift in the cost curves of all the individual firms already in the industry, and the new firm itself would be unable to undercut those already in existence by any substantial amount. In this way the scarcity of one or more factors of production can give a full explanation of the fact that an industry's supply curve is upward sloping. Not only will marginal costs be rising in each individual firm, for reasons we have already studied, but it will be impossible for new firms to undercut those already in the field.

This principle appears most clearly in the case of agricultural production, where land is the scarce factor. After the early stage of abundant fertile land, which has long since been passed in every well-settled community, a large number of comparatively small farms has generally proved to be the organization of maximum efficiency, and the farms normally have their plants fully utilized. Increasing the product of the existing farms means that more laborers must be employed per acre, more capital must be invested in the form of buildings, farm machinery, tools, drainage systems,

etc., more must be spent on fertilizers, and so on. All this means an increased cost per unit of product. On the other hand, there is only a limited amount of idle land available, and this land is of inferior quality to that now used; otherwise it would be already in use. New farms, because of the inferiority of the lands, will from the beginning have to be operated at higher costs than prevail on the existing farms. Since either means of obtaining an increase in agricultural products thus involves increased unit cost, both will be resorted to, and the increase will come partly from the cultivation of additional inferior lands and partly from more intensive cultivation of the lands already in use.

Constant costs. A different supply situation is found in certain lines of industry in which, after a certain price has been attained, it is normally easy for new firms to enter the market with costs no higher, or only a shade higher, than the prevailing price. For an industry to find itself in this situation two essential conditions must be satisfied. (1) The firms composing the industry must have similar cost curves, and it must be possible for new firms, if they enter the industry, to produce under substantially the same conditions of cost. By similarity of cost we mean, more specifically, that the minimum points on the average cost curves of the various firms must be at about the same level. Usually in such an industry the plant of most efficient size is comparatively small. If the most efficient sized plant is very large, the industry is almost certain not to be operating under conditions of pure competition; and even if it is, it becomes difficult and hazardous for new enterprises to enter the field. (2) There must not be heavy reliance upon a scarce factor of production. If there is, the entrance of new firms will raise the general level of costs for all producers so that it will be impossible to expand output except under the stimulus of rising prices.

The separate producer as usual looks upon the prevailing price as a determined thing. Each will furnish that amount of product which equates his marginal cost to the price. No firm at present operating will increase its output unless a higher price is offered, because under perfect competition the supply curve of any individual firm must always be upward sloping. But since we have

assumed that new firms may enter the industry with approximately the same costs as those already in existence, a very slight rise in price will induce a large expansion of output through the entrance of new firms. Conversely a very small drop in price will lead to a heavy shrinkage in output. We have assumed that the firms already in the industry have similar cost curves; if this is the case, competition will have driven the price down to a level where it just covers the average cost (including of course normal profits) for each of the firms in the industry. It will require therefore only a slight drop in price to render the operations of many of the firms unprofitable and thus, in the long run, to drive them out of the industry entirely.

In this as in the preceding case, we assume that new firms will be attracted into any purely competitive industry as soon as an opportunity for profit appears there. Whenever the cost of all existing producers and of any new ones that may enter the industry are approximately the same, the price cannot fall appreciably without heavily reducing the outputs of the industry, nor can it rise appreciably without stimulating an indefinite expansion of output. Each enterprise will have an upward sloping individual supply curve, but the output of the industry as a whole can be enlarged without any increase in the general level of costs. In place of a normal ascending supply curve we have, after a brief initial upward slope, a substantially horizontal supply curve (illustrated by the line ss' in Figure 20).

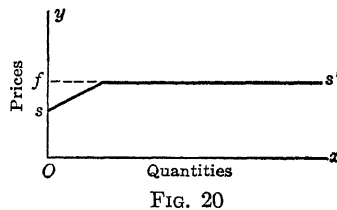


FIG. 20

It may at first thought seem paradoxical that, since the total supply is the sum of the several individual supplies, and since each individual firm has an upward sloping supply curve, we can derive a constant or horizontal total supply curve. The paradox is resolved by recalling that for each individual firm that part of its marginal cost curve that lies above the prevailing price is of no effect, since new firms are prepared to enter the market and supply an indefinite product without going above the prevailing price.

Our total curve is obtained by adding the individual curves such as mm' in Figure 21, but only up to the point p .

Elasticity of supply. It is interesting to note that the horizontal supply curve or the condition of constant costs for the industry as a

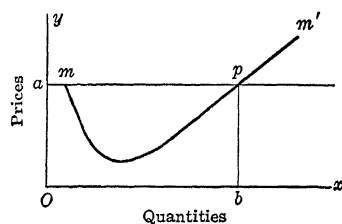


FIG. 21

whole represents a state of perfect elasticity of supply. There is variation in the elasticity of the supply of different commodities, analogous to that in the elasticity of demand. If the quantity of a good offered for sale varies a great deal as compared with variations in price, we have an elastic supply.

Here the supply curve has a gradual slope. If the quantity offered varies only slightly in response to changes in price, we have an inelastic supply. Here the supply curve has a steep slope. In the case of supply however it is not possible to distinguish between less pronounced examples of inelasticity and elasticity by means of the total amount of money involved — as we did with demand. The total amount of money, viewed from the supply side, varies almost always directly with both price and quantity.

In general, within the range of usual prices, the supply of the products of the extractive industries is inelastic, whereas manufactured articles are elastic in supply, though these rules are by no means without exception.

The absolute limit of elastic supply is reached when sellers are ready to offer any quantity at a given price but nothing at all, or only a very much smaller quantity, at any lower price. In other words, the price is the same irrespective of the quantity. The graphical representation of this situation is a straight horizontal supply curve. (See Figure 20.) This condition is characteristic of industries operating at constant cost, as we have just seen.

The absolute limit of inelastic supply is reached when the quantity people are able and willing to offer for sale is the same irrespective of price. Such supply is represented graphically by a vertical straight line. It is scarcely conceivable that this extreme could be reached in connection with pure competition of many sellers.

Since all are subject to rising marginal costs, increase in price would be sure to bring forth increased product from some if not all producers.

No supply curve slopes downward. To round out our analysis of the conditions of supply, we must here observe that there can normally be no such thing as a downward sloping supply curve. Disregarding such exceptional cases as the improvident savages who quit work just as soon as they have received a certain amount of pay, how could a lower price possibly induce sellers to offer larger quantities? The relation between price and quantity offered, in a perfectly competitive market, is such that the latter varies directly with the former, thus giving a supply curve that always slopes upward except when it becomes a horizontal line at the limit of absolute elasticity of supply. The rising and horizontal supply curves go respectively with conditions of increasing and constant costs in the industry as a whole. There may be also, as we shall soon see, a condition of decreasing costs in an industry, but that does not mean that there is a downward sloping supply curve.

The law of supply. From the foregoing analysis the following generalizations — which we may call *the law of supply* — follow quite obviously. *Immediately, or in the short run, in a given market at a given time, the quantity of any good which people are ready to offer for sale varies directly with the price. In the long run, the same principle holds, except that at the limit of absolute elasticity of supply the quantity will vary without limit if price departs from a certain magnitude.* This law is readily illustrated graphically by upward sloping and horizontal supply curves.

Long run and short run again. Of course the practical interests of business men and of the consumers generally are far more concerned with long run supply conditions than with the seller's short run sales problems. And the discussion in this chapter has applied primarily to the long run. If for example we are studying the variations in the output of cotton that occur within a single year in response to variations in price, it is obviously not realistic to assume that large numbers of new firms will within that period be attracted into the industry by an increasing price or that existing firms will be driven out of production. For short run considerations

we will achieve a much closer approximation to reality by considering the number of enterprises as fixed, giving us a total supply curve based on the individual supply curves of all existing firms. It is only if we are concerned with long run tendencies that we may legitimately assume that high profits will attract new producers and very low profits will drive existing firms out of business.

Though this point is fairly clear in its application to all sorts of production, its significance becomes especially pronounced when we are considering industrial rather than agricultural enterprise. Most agricultural production is carried on with highly durable equipment, the purchase of which has given rise (as we have seen) to fixed costs. Price would fall to a level well below the minimum average costs of many of the producers in the industry without, in the short run, causing them to cease production. Thus a decline in price will lead to a much smaller reduction of output in the short run than in the long run, just as a rise in price will not have its full effect for a considerable period of time. In our discussion of increasing costs and constant costs we have found that the shape of the total supply curve will depend, among other things, upon the ease with which new firms can enter the field. Clearly therefore these supply curves refer to the long run rather than the short run.

Further analysis of the nature of supply. Before going any further, it is advisable to admit the limitations of the technique we have been developing, at the same time stressing its importance as an aid toward understanding. The supply curves we have been considering are graphical representations of comparatively simple supply schedules. These schedules show quantities people are able and willing to offer at various prices — subject to certain qualifications which have not been openly expressed. None of these schedules is valid unless the prices of other goods and services either are not subject to change in connection with change in the price of the good in question or, if so subject to change, bear a known relation to the price or quantity of the good under consideration. Thus a supply schedule for shoes may be drawn up on the assumption that the price of leather and the price of labor (wages) vary only as affected, in a known relationship, by possible changes in the production of shoes. A supply schedule for railroad

ties may be drawn up on the assumption that there are no changes in the price of lumber for general construction purposes. The price of leather in the first case may really be affected by changes in its use for other purposes than shoes, for example, belts, saddles, etc.; and more important still, the price of leather may be affected by changes in the cattle industry which have no relation to the production of shoes. Wages in the shoe industry are subject to variation as a result of changes in other industries, which may or may not correspond with those in the shoe industry. Likewise the number of railroad ties which may be delivered at a given point depends not only on the price of the ties themselves but also on other prices — wages again — and the prices of other uses to which the wood may be put. It is obvious that there is no indication of responses to all these possible contingencies in a simple supply schedule or curve. In the most profound sense, the supply of any good is, like demand, a schedule of schedules, which takes into account not only all possible variations in the price of the good in question but also all possible variations in all other prices.

Decreasing cost industries. We have warned ourselves in earlier paragraphs of this chapter against the fallacy of a downward sloping supply curve. Since a fall in price will normally induce existing firms to restrict output and could not conceivably induce new firms to enter an industry, it is quite inconceivable that a fall in price should induce an increase in output except in a very abnormal situation. This does not mean however that there cannot be an industry so constituted that as its output increases the general level of costs will be lower. Specifically if an industry is prosperous so that new firms are attracted into it and output is steadily expanded, it is possible that the very expansion of the industry will have the effect of lowering average and marginal costs for the constituent firms. In such an industry we still could not say that there was a downward sloping supply curve, since a decrease in price would certainly not lead to an increase in output. What we could say however is that if (perhaps through an initial increase in price) output is expanded and costs thereby lowered, the industry could be in equilibrium after the expansion had occurred with a lower price than before.

There are two possible ways in which the growth of an industry through the addition of new firms might have this favorable effect upon costs. First we may consider a condition which is the exact reverse of that which gave rise to increasing costs; that is, a condition in which some element of cost in an industry becomes smaller as the industry expands. Such instances are by no means impossible to find. For instance the rapid expansion of agriculture in the Mississippi Valley eventually reached the point at which it became practical to use complicated machinery — harvesters, reapers, steam threshers — in place of the simpler tools which were appropriate to small farming. Unit costs were reduced, leading to successive downward shifts of the cost schedules of individual farmers. Under such conditions it is clear that as the output of the industry increases it can afford to sell that output at a lower and lower price.

The second explanation, frequently concurrent with the first, is that as an industry grows its organization may so improve and the division of function within the industry may so far advance that it can produce at lower and lower costs. As the British cotton industry in Lancashire expanded in the nineteenth and early twentieth century, this process was observable. More and more enterprises entered the field. The degree of specialization among them became greater and greater; firms would restrict themselves solely to the spinning of yarn and others to weaving, and in a still later stage an individual firm would restrict its weaving to certain grades and qualities of cloth. As this process continued, each firm was concentrating more and more on a single specialty and was thus able to produce this particular specialty more and more cheaply. The economy occurred not so much through the addition of new firms as because the improved organization of the industry permitted reductions in price to accompany growth in output.

If now we should undertake to represent this situation graphically, we could draw a downward sloping curve to show for each output the price which would be sufficient to induce the industry to continue producing that output. The reader will be on his guard against thinking this is a supply curve of the industry. Even in such an industry a rising price will stimulate both an increase in output by existing firms and the entrance of new firms into the

field. Conversely a fall in price will induce existing firms to diminish their output and will drive some firms out of the industry entirely. Nevertheless it will be true in such an industry that after an expansion of output the industry could be in equilibrium at lower prices than before, because costs will have declined in consequence of the increase in output.

The business man's interest in the supply side of the market:

In general. In Chapter IX the reader was reminded of the keen interest manifested by modern business men in the demand for their products and of their endeavors to learn all they can about demand and to influence it to their own advantage. To the business man, demand represents the reaction of his customers to his product — the buyer's side of his market. But he is no less interested in the seller's side; that is, in the reactions of other business men, his competitors, not only in the same line of production but in other lines which may compete with his product for the money of the consumers. Business men are thus interested in average and marginal costs of production, in capital charges and overhead, in wage scales and costs of materials, in stocks of merchandise on hand, in the progress of goods in process of manufacture, in freight-charges, and so on. And all these matters interest them, not only in connection with their own enterprises, but with respect to industry as a whole. In short, they desire to know all they can about production and supply, realizing that these are factors upon which depend the prices they can charge for their products and the quantities which they can sell. In recent years well-organized efforts are being made to ascertain the significant facts bearing upon the supply side of the market of all important products.

Measurement of potential supply. Supply has been defined as the schedule of the respective quantities of a good which people are ready to offer for sale at all possible prices. In this technical sense supply is not capable of measurement. It depends upon the decisions of the body of sellers who are offering any particular commodity. However these decisions are in the long run predicated by certain definitely measurable conditions surrounding any particular commodity. The most important of these conditions is the *potential supply*, or *stock*, of the commodity which is available

for sale. This potential supply is the resultant of two factors; *viz.*, current production, and stocks of the commodity produced in past periods but still unsold, known as *stocks on hand*.

Current production may be measured in physical units, such as barrels of crude petroleum, bushels of wheat, board feet of lumber, etc. It is sometimes more suggestive to present the changes in current production by means of index numbers expressing the volume of current production of some other period. Index numbers therefore indicate by what percentage current production is greater or less than the production of some selected past period. Similarly stocks on hand are measured both in terms of physical quantities and by index numbers. But even these data fail to present the complete picture. Properly to interpret potential supply, it is necessary to know whether stocks on hand are larger than normal and whether current production is running above or below normal.

Normal and residual changes. To permit the presentation of these data clearly, statistical technique has developed the concepts of long time movement or secular trend, seasonal variation, cyclical fluctuations, and residual fluctuations. These concepts make possible definite conceptions of what is normal.

The growth of population and the increase in productive output, with the resulting elevation in standards of living, provides the basis for the statistical assumption that the long time movement in the production of any commodity will be continued. The recurrence of seasons with their influences on production provides the bases for the assumption that these seasonal variations will persist in marked degree as long as the influences of weather on production continue. Thus combining the expected effects of long time movement and seasonal variation on the production of a particular commodity, we derive a measurable concept of normal. It then becomes possible to discover whether current production at any particular time is above or below this normal.

By the use of these devices the seller is enabled to know whether production is outrunning consumption. If production is exceeding normal expectations and stocks on hand are increasing, potential supply is increasing, and decreases in price threaten. If on

the other hand, production is above normal but stocks on hand are declining, consumption has exceeded past expectations, and prices can be expected to remain stable or perhaps even to increase. For example the following figures show a condition of increasing production, increasing stocks, and falling price in the gasoline market (the quantities being monthly averages for the respective years, expressed in barrels):¹

	1921	1928	1938*
Production, raw at refineries	10,225,000	31,432,000	45,599,000
Consumption	8,960,000	27,403,000	42,749,000
Exports	1,058,000	4,302,000	3,373,000
Stocks at end of month	15,018,000	33,759,000	76,061,000
Price, wholesale at New York, per gal.	\$0.261	\$0.174	\$0.128

* Monthly average for first nine months of 1938.

Indexes of potential supply. While the seller is primarily interested in his own particular commodity, he is not immune from

INDEX OF INDUSTRIAL PRODUCTION²

Product	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.
Manufactured (total)	76	75	75	73	73	74	82	87	89	95 ^p
Iron and steel	52	50	49	50	47	46	62	70	76	90
Textiles	75	79 ^r	81	74	77	87	97	110	103	100 ^p
Food products:										
Slaughtering and meat packing	92	86	83	84	84	81	83	89	98	95
Wheat flour	86	89	90	90	86	98	98	87	88	91
Sugar meltings	99	83	63	77	81	78	95	86	106	103
Newsprint consumption	132	129	126	125	125	122	124	123	130	131
Automobiles	65	61	54	54	49	46	43	45	46	84
Leather and products	94	102	104	104	105	90	103	108 ^r	102	99 ^p
Cement	59	57	65	67	69	69	71	67	69	80
Nonferrous metals:										
Tin deliveries	91	85	77	71	69	73	68	65	71	81
Zinc	98	90	87	80	79	70	69	74	75	80
Lead	69	67	64	74	60	64	54	46	50	50
Petroleum refining	200	194	191	196	197	192	199	203	206	—
Rubber tires and tubes	67	57	62	64	65	72	82	91	96	101
Tobacco products	157	157	160	159	163	154	154	161	160	150
Mineral production (total)	108	103	103	101	91	92	93	95	97	100 ^p

^p Preliminary.

^r Revised.

¹ *Survey of Current Business*, December, 1938, p. 45.

² *Federal Reserve Bulletin*, December, 1938, p. 1077.

developments throughout the whole range of business. There are available for his purposes elaborate indexes of production, summarized by large classes, of which the foregoing are samples. The base (100) is the average of the period 1923-1925. There has been an adjustment for seasonal variations.

Decisions of sellers are predicated, not only on the potential supply that may be thrown on the market in the immediate future, but also on the supply in the more extended future. Sellers accordingly look to more general indicators than those of current production and stocks on hand. Thus the expansion of bank deposits, the import or export of gold, the policy of the Federal Reserve Banks, the activity in security markets, the productivity of labor, and the policy of labor organizations represent influences acting upon interest cost, wages, prices of raw materials, freight rates, etc.

Seeking control of supply. The business man may carry his interest in supply a step further still and, as in the case of demand, cast about for the means to control supply to his own ends. Except in the case of the few great combinations, the individual business man generally realizes that he can alone do little to affect the supply or the general cost conditions of his product, being, as it is, too small a fraction of the whole. But he sees that through common action by all competitors in his line, there might emerge agreements as to prices, division of territory, uniformity in sales practices, interchange of information as to costs, stocks on hand, orders for new business, etc., whereby the supply might be at least in some degree subject to control to the advantage of all producers. It is here that the business man sets his feet on the path toward monopoly and inclines to run up against public opinion and the law through action "in restraint of trade."

For centuries it has been a common practice of governments to reduce the supply of certain products by means of a protective tariff, which prevents or restricts the importation of certain foreign goods. The important problem of the protective tariff will be examined in a later chapter of this book.

Attempts to control supply in the interest of certain favored producers have frequently been made by governments. For example

the Brazilian government since 1931 has destroyed millions of bags of coffee in an attempt to support the price through reduction in supply. The United States government has made strenuous efforts thus to aid the farmers. From 1930 to 1933 the Federal Farm Board sought to raise the prices of wheat and cotton by purchasing large quantities and taking them off the market. The "New Deal" program, starting in 1933, undertook, through government leasing of farm lands and "benefit payments" to farmers, to reduce the acreage devoted to cotton and certain other crops. More extreme measures were the plowing under of growing cotton and the slaughter of young pigs. Finally a law in 1934 set a definite limit of ten million bales to the cotton crop of 1934-35, with a tax of 50 per cent on the excess. These measures will receive further attention at later points in this book.

EXERCISES

1. Assume the following relation between output and marginal cost for the production of potatoes on four different farms:

<i>Marginal cost</i>	<i>Farm A</i>	<i>Output (in bu.)</i>			<i>Farm D</i>
		<i>Farm B</i>	<i>Farm C</i>		
\$0.20	1,000	400	700	600	
.30	2,000	800	1,200	1,200	
.50	3,000	1,200	1,700	1,800	
.80	4,000	1,600	2,200	2,400	
1.20	5,000	2,000	2,700	3,000	
1.80	6,000	2,400	3,200	3,500	
2.75	7,000	2,800	3,700	4,000	

- Draw up a total supply schedule to include these four potato producers.
- Construct the supply curve corresponding to this supply schedule.
- At \$1.00 per bushel, what quantity of potatoes would be produced and offered for sale?

2. Describe the essential features of a supply schedule. What factors may cause a change in supply from one time to another? Draw a simple diagram to illustrate an increase in supply; a decrease in supply.

XIII

PRICE UNDER PURE COMPETITION

Pure and imperfect competition. The prices of wealth and services are determined in any market by the action of forces coming from buyers on the one side and sellers on the other side. The forces on the buyer's side operate through the demand for the good in question.

In an imperfectly competitive market, the actions of sellers depend upon ruling price, elasticity of demand, and the effectiveness of devices for encouraging sales. Under these circumstances the analysis of the forces which bear upon the market from the seller's side is somewhat complicated. But under conditions of pure competition we are able to represent these forces in terms of the supply of the good, which greatly simplifies our problem. The present chapter is confined to the problem of price determination under conditions of pure competition.

Determination of price and quantity exchanged. Equilibrium: **Price determined by demand and supply.** Demand and supply, hitherto separately examined as representing respectively the buyer's side and the seller's side of the market, may now be brought together. In order to have the assistance of a concrete example, let us assume that we are investigating the price of fresh-laid eggs in the retail market of a small sized city, which we call T. We assume that on a certain day, October 15, 1939, there were no changes in other prices, but that the price of eggs at the beginning of the day was somewhat uncertain. The demand for eggs and the supply of eggs were as shown in the schedules on page 271. The relations between the quantities and the prices in these schedules conform to established laws of demand and supply. The exact numerical quantities are of course not significant; they are used only for purposes of illustration.

Now let us plot on one diagram two curves corresponding respectively to these two schedules. The result is shown in Figure 22; *dd* is the demand curve; *ss*, the supply curve.

RETAIL EGG MARKET OF T, OCTOBER 15, 1939

<i>Demand</i>		<i>Supply</i>	
PRICES	QUANTITIES PEOPLE WERE READY TO BUY	PRICES	QUANTITIES PEOPLE WERE READY TO OFFER
90 cents	160 dozen	90 cents	375 dozen
85 cents	170 dozen	85 cents	360 dozen
80 cents	185 dozen	80 cents	340 dozen
75 cents	200 dozen	75 cents	320 dozen
70 cents	220 dozen	70 cents	300 dozen
65 cents	240 dozen	65 cents	280 dozen
60 cents	260 dozen	60 cents	260 dozen
55 cents	280 dozen	55 cents	230 dozen
50 cents	300 dozen	50 cents	200 dozen
45 cents	320 dozen	45 cents	160 dozen
40 cents	340 dozen	40 cents	120 dozen
35 cents	360 dozen	35 cents	80 dozen
30 cents	380 dozen	30 cents	40 dozen

The curves of demand and supply intersect at the point p . The height of this point, pb , represents the same price, and its distance from Oy the same quantity, on both curves. The point p , as a point

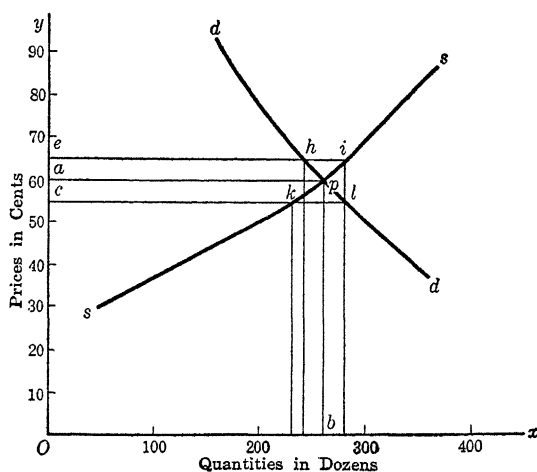


FIG. 22

on the curve dd , indicates that at a price of 60 cents people in the city of T were able and willing to buy 260 dozen eggs; as a point on the curve ss , p indicates that at a price of 60 cents other people in

the same city were able and willing to sell 260 dozen eggs. At the price of 60 cents quantities demanded and offered were the same.

Since we are assuming a market of pure competition — in which there are many independent buyers and sellers, each familiar with the conditions of demand and supply of eggs on this particular day — this is *the price* of eggs in this market on this day. The price could not be anything else. Any seller who would demand more than this, say 65 cents a dozen, would find nobody willing to buy from him. Rather than be left with his stock unsold on his hands, he would reduce his price. A seller, on the other hand, who offered to sell eggs for less than 60 cents, say 55 cents, would soon find himself swamped by a rush of eager buyers; he would see that he could easily sell all his stock for more than 55 cents, and he would raise his price. We reach the same conclusion on regarding the matter from the buyer's standpoint. A purchaser who offered to buy any considerable quantity at more than 60 cents a dozen would find all the sellers eager to sell to him and would soon see that he could get all the eggs he wanted at a lower price. He would therefore lower his price. Any buyer, on the other hand, who tried to get eggs for less than 60 cents would find no one willing to sell to him and would have to raise his price or go without. No considerable buying and selling could therefore take place at any price except 60 cents, so long as our curves accurately represent the conditions of demand and supply and buyers and sellers in general are aware of what is going on in the egg market.

Determination of the quantity exchanged. Clearing the market.

It will be evident that when demand and supply determine the price they also determine the quantity exchanged. Graphically, the intersection of the curves at p (Figure 22) determines the quantity exchanged at 260 dozen, as well as the price at 60 cents. At the price thus determined all sellers of eggs are able to dispose of the entire quantity of eggs which they are willing to sell (at that price) and all buyers are able to get as many eggs as they wish to buy (at that price). No seller in our example who is willing to take 60 cents a dozen will be left with an unsold stock, and no buyer who is willing to pay 60 cents will fail to get what he wants. In technical terms,

this price "clears the market." This is not true of any other sum. If, assuming for the moment the impossible, the price were 65 cents, all buyers would obtain all they wanted at that price; *i.e.*, 240 dozen; but sellers, willing to dispose of 280 dozen, could sell only 240 dozen, because buyers would take no more at that price. The difference, 40 dozen (represented by hi , the difference between ei and eh), would be left on their hands. On the other hand, consider a price of 55 cents, again assuming the impossible. Sellers would be able to dispose of all they had for sale at that price; *i.e.*, 230 dozen. But buyers, although willing to buy 280 dozen, would be able to get only 230 dozen, since the sellers would part with no more. The difference, 50 dozen (represented by kl , the difference between cl and ck), stands for the unsatisfied requirements of the buyers. Such differences cannot exist at the *price pb*, determined by demand and supply.

A free market. We may arrive at this same conclusion in slightly more theoretical terms by reference to the definition given in Chapter IX of a free market. In a free market a buyer or a seller must always have the right to determine either quantity or price, but not both. Consider in the above example the case in which the price was set at 55 cents. At that price buyers wish to buy 50 dozen more eggs than they could actually secure. If this condition persisted, we could not say that they were dealing in a free market, because producers would be imposing rationing upon them. If the buyers were thus deprived of the right to buy the amount which they chose, they would have to be granted the right to bid up the price. Thus in any market which actually operates as a free market and where there is no rationing on either side, the only price which can obtain will be that which will induce sellers to sell the exact number of units that buyers are induced to buy.

The laws of demand and supply. So far in our investigation of the egg market, buyers and sellers in general have been assumed to be well informed and prompt in action. Such is not always the case. The result is a slowing up of the interaction of the forces of demand and supply. But just as murder will out, so facts regarding the egg market percolate through the community and tend ultimately to have the same effect as they would immediately under more active

conditions. On the first day we study the egg market, the price of 65 cents a dozen may be carried over from a previous day; but on October 15 it leaves at least some egg dealers with a surplus on hand which they would have sold at this price if they had had the chance. These sellers may not worry about it because they think it is just an off day, and they hope to make up for it tomorrow. But if other days follow with similar conditions of demand and supply and similar results, someone will eventually react and reduce the price of eggs. His motive is to get rid of his surplus. As soon as any sellers reduce their prices, others, offering the same kind of eggs, begin to reduce their prices also. They may have been lucky enough to sell all they were willing and able at the price of 65 cents; but as soon as competitors reduce their prices, they are in danger of losing their business. They may not indeed lose all of it, but they may lose enough to affect them seriously if they do not meet the reduction in price. Thus if conditions of demand and supply remain the same, the price of eggs will drop toward 60 cents.

On the other hand, on the first day we study this market the price of 50 cents a dozen may have been carried over from a previous day. As a result, some people who want to buy eggs do not get them. This may not have any effect on the price that day, because sellers do not immediately notice the fact. Soon however someone will notice it and put up the price. In this case he knows that he need not fear the lower price of competitors, because they are not able and willing to sell all that is asked at that price. As soon as one seller is successful in selling at a higher price, others follow suit, seeing the opportunity for additional profit. Thus in this case, if conditions of demand and supply remain the same, the price of eggs will rise toward 60 cents.

Observation of this process of adjustment led men long ago to formulate what are commonly known as *the laws of demand and supply*. There are many laws of demand and supply. We have already considered separately a law of demand and a law of supply. When people refer to "the laws of demand and supply," they generally have in mind the joint action of forces on both sides of the market, which may be described as follows :

I. If the quantity of any good which people are ready to buy at the prevailing price is greater than the quantity offered at that price, the price will rise.

II. If the quantity of any good which people are ready to buy at the prevailing price is less than the quantity offered at that price, the price will fall.

III. At equilibrium, the quantity of any good which people are ready to buy will be equal to the quantity offered.

In all economics there is nothing more important than these three principles. They lend themselves to simple graphical illustration, as in Figure 23.

(1) If the price were Oa , we would have af greater than ae , and the competition of buyers would cause the price to rise. (2) If the price were Oc , we would have ch greater than cg , and the competition of sellers would cause the price to fall. (3) At a price of Ob , the length of bp indicates a condition of equilibrium.

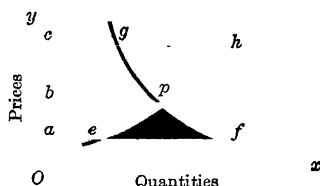


FIG. 23

Market price and equilibrium

price. In any active market which is perfectly competitive there is a tendency toward uniformity of price. There may be scattered sales at various prices, but at any given time most transactions take place at the same price, or thereabout. It is convenient to use the term "market price" to indicate the price at which the bulk of exchanges in any given commodity are taking place at a given time in a particular market. This may or may not be the same as the "equilibrium price"; i.e., the price which clears the market, though market price is always tending in this direction. In measuring the depth of the ocean, for example, one cannot take account of all the waves nor the troughs between waves; it is sufficient to consider the general level at which the waves fluctuate. So in a general description of price behavior we cannot take account of all the minor irregularities of prices. In further analysis therefore we assume that market price and equilibrium price are the same. This is not always accurate, but it is accurate enough for general understanding.

The law of price. We arrive thus at the following *law of price*. *In any purely competitive market the price of any good is determined by the two forces of demand and supply, at that point at which the quantity which people are ready to buy is equal to the quantity which people are ready to offer for sale, and that the quantity of the good exchanged is simultaneously determined by the same forces.*

These are the determining forces, and the sole determinants. This may at first seem a sweeping saying. Is not the price of eggs affected by the activity of the hens in laying? Does not the rainfall affect the price of wheat? Are not the prices of many articles affected by changes in fashion? Was not the price of liver raised by general dissemination of knowledge as to its beneficial effects in human diet? And are we not always being told that the "high cost of living" is due to "inflation of the currency"? Such pertinent questions might be multiplied, but their obvious affirmative answers do not imply that here is a host of additional causes affecting competitive price. All these circumstances, and others, do influence prices, but they do so only indirectly. They act only by first affecting either demand or supply. A drought reduces the wheat harvest; the supply is diminished, and the price of wheat goes up. Knowledge of its qualities increased the demand for liver and so raised its price. There is no side door through which any influence can get at price; price may be approached only through one of the specified main entrances, remembering always that we are here concerned only with the competitive prices of goods subject to conditions of increasing cost.

Alternative concepts of demand and supply: **Two possible meanings of demand.** Before proceeding further in the investigation of the forces determining price and quantity exchanged, we shall find it advisable to pause here for a somewhat critical consideration of our definitions and a closer scrutiny of the nature of demand and supply. The whole investigation of the laws of price is in danger of being confused by the fact that the term, demand, is used with two different meanings in popular speech and writing and, it must be admitted, in much scientific writing also. A clear understanding of these two meanings will enable the student to avoid much confusion in his later reading and discussion. For example we hear

“there is a heavy demand for wheat”; “the demand for high shoes is decreasing rapidly”; “the demand for automobiles is unusually strong this spring”; etc. In all such expressions it is evident that the word, demand, refers to the whole *demand schedule* (or curve). The meaning is that at any given price the quantity that buyers would choose to take is large or small, greater or less, etc. In Figure 24, the curve $d'd'$ represents a heavy demand as compared with the

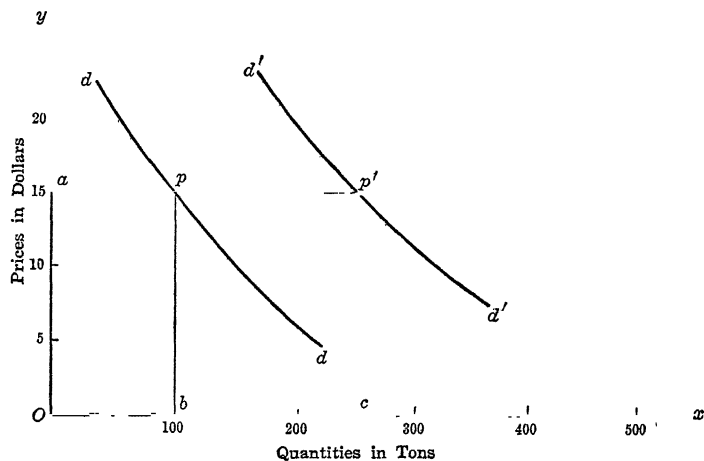


FIG. 24

curve dd , since whatever price is chosen the quantity that would be taken according to $d'd'$ is greater than according to dd . Thus if dd represents the demand for coal at a given time, a price of \$15 would on this day induce buyers to take 100 tons; if on some other day the demand were as shown by the curve $d'd'$, this means that buyers would on that day take 250 tons at \$15 a ton. The same notion of demand is likewise in mind when such expressions as these are used: “the price of anything depends on demand and supply”; “the greater the demand, the higher the price”; etc.

The foregoing usage evidently agrees with the definition of demand which we have accepted.¹ On the other hand, one meets such statements as these: “an increase in price causes a decrease in demand”; or “the higher the price, the smaller the demand.” Now it is evident that these statements are in direct conflict with the last

¹ See Chapter VIII.

quotation in the previous paragraph. Something must be wrong when people say "the greater the demand, the higher the price" and, in the next breath, "the higher the price, the smaller the demand." The trouble comes from shifting the meaning of the word, demand. In the second expression demand does not mean the whole demand schedule or curve. It means rather *the quantity people are ready to buy at some particular price*. Refer again to Figure 24, and note that, according to the curve *dd*, at a price of \$15 buyers would take 100 tons of coal. This quantity, 100 tons, is the *demand for coal at \$15*, when the word is used in this second sense. It is represented by the line *ap* or the line *Ob*. Suppose on the day the curve *dd* is in effect, someone says the demand for coal would be less if the price were raised above \$15, or the demand would be greater if the price were less. He does not mean a change in the whole demand schedule or curve, but rather a shift from one to another of a series of simultaneous possibilities. The ability and willingness to buy any other quantity at any other price, as shown by the curve, exists at the same time as the ability and willingness to buy 100 tons at \$15. In graphical terms this talk of a decrease or increase in demand is described by a movement along the curve, locating a different point, whose distance from the *y* axis is greater or smaller than before. When used in this sense we can never think of demand except in reference to a *particular price*, stated or implied. We cannot speak simply of the demand for cotton. We must say: the demand for cotton at fifteen cents a pound, or at some other price.

Choice of a definition of demand. Now it is idle to argue about definitions; and this presentation of the two senses in which the term, demand, is used is not offered as an introduction to a battle of words. As has been previously stated,¹ there is no absolute test of the correctness of a definition. Either of the two possible meanings of the term, demand, might be chosen. Both are about equally favored in the loose usage of popular speech and writing. To that extent we violate popular usage when we choose to make exclusive use of either meaning; but that is inevitable. We are free then to select the meaning which will best serve the purpose of the scientific investigation of economic laws. The central problem in economics

¹ See Chapter I.

is concerned with the laws which determine value and price. In the study of this problem, while it would be possible to make fairly satisfactory progress with either definition of demand, the advantage is decidedly in favor of the first meaning. We have therefore accepted the definition: *The demand for any good is a schedule of the respective quantities of that good which people are ready to buy at all possible prices.* The justification of this choice will appear as the reader proceeds.

Two possible meanings of supply. Choice of a definition. The term, supply, is commonly used in two senses, corresponding exactly to the two meanings which are given to demand. When people say "the supply of cotton is short this year," "the wheat supply is greater than ever before," etc., they are using the term as the whole *schedule of quantities that people are ready to offer at various prices.* They mean, for example, that, at whatever price bid, the quantity of cotton which sellers are ready to offer is small, relative to the quantities they are ordinarily ready to offer. The term is used also in this sense, when it is said: "the price of anything depends on demand and supply"; "the greater the supply, the lower the price"; etc.

This usage is obviously in harmony with the definition of supply which we have adopted. However such expressions as these will frequently occur: "an increase in price causes an increase in supply"; "the supply depends on the price"; "the lower the price, the smaller the supply." There is here the same confusion as results from the two meanings of demand. For example the last quotation above is clearly in flat contradiction with the last quotation of the previous paragraph. As clearly, the trouble is due to shifting the meaning of the term, supply. In the sentences quoted in this paragraph, "supply" does not mean a schedule of quantities that would be offered at various prices; it means *the quantity people are ready to offer at some particular price.* Referring to Figure 25, it will be seen that, at a price of 70 cents, 7,000 railroad ties would presumably be offered. In the second sense of the term, supply, it is this particular quantity that is the *supply of railroad ties at 70 cents.* It is represented by the line *ap* or the line *Ob*. When one speaks of an increase or decrease in supply in this sense, he means, not a

change in the whole supply schedule, but a shift from one to another of its simultaneous possibilities. The readiness to offer any other quantity at any other price, as shown by the supply curve, exists at the same time as the readiness to offer 7,000 ties at 70 cents. In graphical terms this is described by a movement along the curve, thus locating a different point whose distance from the y axis is greater or smaller than before. Supply, in this sense, is always in reference to a particular price. We cannot conceive of the supply of

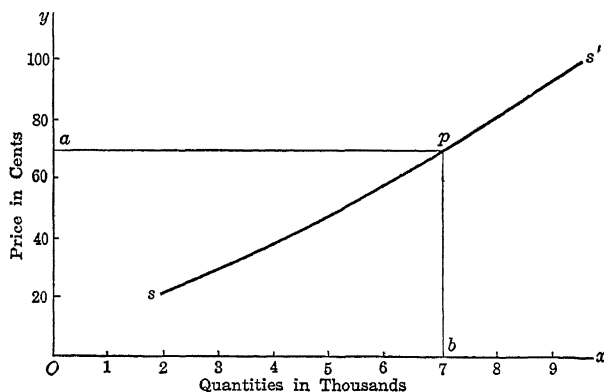


FIG. 25

wheat in general. It must always be the supply at \$0.93 a bushel, or some other price.

All of this corresponds evidently to the two meanings of the term, demand. Here also it is true that fairly satisfactory progress might be made with either definition, but that the advantage is with that which regards supply as the whole schedule rather than a particular quantity offered. Hence we have adopted this definition: *The supply of any good is a schedule of the respective quantities of that good which people are ready to offer for sale at all possible prices.*

The importance of consistency. While satisfactory results might come from the consistent employment of either pair of possible definitions of demand and similarly of supply, nothing but confusion can result from a vague use or from a mingling of the two meanings which might be attached to each of these terms, twin sins of which it must be confessed that some very good economists have

been occasionally guilty and which run riot in popular speech and writing. Equally essential is it that the definitions chosen for demand and supply be in harmony with each other. We cannot define demand, for example, as a schedule of quantities that would be bought, while defining supply as the quantity that would be offered at a particular price. We have chosen, for reasons stated, which will become more apparent as our study proceeds, the schedule sense for both terms. These definitions will be adhered to consistently throughout this book. The reader who goes on with his study of economics is sure to meet writers whose usage is different. If he has mastered the lesson here presented, he will have little difficulty with those writers who have chosen and used consistently the alternative definitions. He will have more difficulty in understanding those writers whose usage is vague or who mix the two definitions, but even here his task will be lightened if he has a clear notion of what the two possible meanings are and how they may be used.

The causes of price change: In general. Our excursion into the theory of price has thus far shown us how competitive prices of goods subject to conditions of increasing costs are determined, as of a particular time, by demand and supply. Prices in the practical world of today are however seldom stationary but are generally subject to change, more or less frequent. It is not enough therefore to know how a certain competitive price may be established in a static market. We must also understand the mechanism by which a price may rise and fall under the dynamic conditions of modern life.

From our previous conclusion that prices of the sort we are now considering are determined solely by the action of demand and supply it follows that changes in price may result only from changes in demand or in supply or in both. A glance at the diagram in Figure 22 (page 271) will determine that no other equilibrium price than pb is possible with those particular curves of demand and supply, since p is the only point at which those curves can intersect. A different point of intersection could be obtained only by a shift in one or both of the curves. With the path thus indicated our present inquiry is quite simple. We shall find the graphical aids especially useful in this part of our investigation.

A change in demand. Let us consider an increase in demand. This means that buyers have become more eager. In terms of the definition of demand, it means that at any given price the quantity that people are able and willing to buy is greater. This may be visualized by use of the illustration of the retail egg market of T (see Figure 22). We there assumed a certain demand and a certain supply resulting in a price of sixty cents on October 15. As cold weather comes on, the people of T incline to consume more eggs; in other words, the demand increases. Let us suppose that the demand on December 15 would be correctly represented by the following schedule:

DEMAND FOR EGGS, RETAIL MARKET OF T, DECEMBER 15, 1939

<i>Prices</i>	<i>Quantities</i>
90 cents	300 dozen
85 cents	320 dozen
80 cents	340 dozen
75 cents	360 dozen
70 cents	380 dozen
65 cents	400 dozen
60 cents	420 dozen
55 cents	440 dozen
50 cents	470 dozen
45 cents	500 dozen
40 cents	540 dozen
35 cents	580 dozen
30 cents	630 dozen

The reader is again reminded that the exact quantities in such a schedule are not significant but are used only for purposes of illustration. What we have is a new schedule in which, for each price, the quantity people are ready to buy is greater than it was before.

In Figure 26 the demand and supply curves as of October 15 are repeated (see Figure 22), and there is added the curve $d'd'$ representing the demand on December 15, as plotted from the schedule of the demand on that date. It will be evident that an increase in demand is represented graphically by a new curve, each point on which is farther away from the y axis than each corresponding point on the old curve; in other words, the demand curve has been shifted to the right. The increased distances from the y axis indi-

cate that now, at each price, the quantity people are ready to buy is greater than before. That is exactly what is meant by an increase in demand. It is what people mean when they say: "the demand for eggs has increased since October"; "there is a strong demand for eggs today"; and such expressions.

What are the effects of the new demand? Let us assume for the sake of simplicity that there has been no change in the conditions of supply in the egg market of T. The curve ss (Figure 26), representing the supply on December 15, is then the same as for October 15.

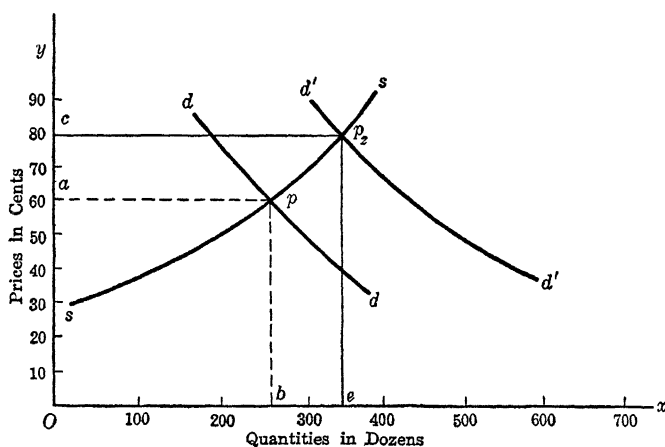


FIG. 26

The curve $d'd'$ represents the demand on December 15. This new demand curve intersects the supply curve at the point p_2 , farther to the right and farther up than the point p , which determined the price and the quantity exchanged on the previous date. The price is now 80 cents, and the quantity of eggs exchanged is 340 dozen, an increase in both magnitudes since the earlier date. We conclude then that under conditions such as these an increase in demand tends to cause an increase in price and an increase in the quantity exchanged.

By employing similar reasoning the reader may verify for himself the converse; *i.e.*, that a decrease in demand tends to cause a decrease in price and in the quantity exchanged.

These conclusions may be readily verified by experience. For

example in the wheat corner in 1917 the extraordinary price of \$3.25 was recognized at the time as the result of the great demand for wheat represented by the purchases of the British government and other foreign governments. The high prices of farm land in the Mississippi valley in 1920 were the result of an increase in demand for farms on the part of those who foresaw a glorious prospect of profits from the prevailing high prices of grain. These are some of the more spectacular examples, but they should not be permitted to obscure the significance of the constant influence upon price of changes in demand operating in less conspicuous ways. The reader will readily discern the influence of a decline in demand in the following quotation: "While spindle activity in cotton manufacturing has fluctuated rather narrowly between 30,000 and 33,000 active spindles from 1923 to 1928, billings of finished cotton goods fell from a monthly average of 95,098,000 yards in 1923 to a monthly average of 75,100,000 yards in 1928. The composite index of cotton prices fell from 213 to 163 on the base of 1911-1913 = 100."¹

A change in supply. Price and quantity exchanged are affected also by changes in supply. An increase in supply means that sellers are more eager to part with their goods, that at any given price the quantity offered would be more than before. For example the months of October to January are a period during which the average hen is at a low stage in her egg-laying activity. From March to June things are going much more briskly. An investigation carried out by the Connecticut Agricultural College showed that the average production per hen at this season was nearly three times what it was at the low stage in October to January. This increased production has an effect on the supply of eggs, which, as everyone knows, is normally much greater in the spring and summer than in the late fall and winter. Graphically, the change would be represented by a new supply curve farther to the right than the previous one and therefore cutting the demand curve at a point lower down and farther to the right, indicating a lower price and a greater quantity exchanged. We reach the general conclusion that, under such conditions as are here assumed, an increase in supply tends to

¹ United States Bureau of Foreign and Domestic Commerce, *Survey of Current Business*, August, 1929, p. 31.

cause a decrease in price and an increase in the quantity exchanged.

By precisely analogous reasoning we arrive at the generalization that, other things remaining equal, a decrease in supply causes a rise in price and a decrease in the quantity exchanged.

The harvests furnish a good illustration of these generalizations. During the growing season the wheat market receives frequent reports as to the area planted to wheat, the progress of the crop, weather, rainfall, and other conditions affecting the future harvest. On the strength of these reports, sellers determine what quantities they can offer at various prices; that is, they determine the supply. A succession of favorable reports will thus increase the supply on the market, whereupon the price will fall and the amount sold will increase.

Crop failures or adverse reports on crop conditions always reduce the supply, and, other things being equal, the result is a rise in price and a falling off in sales. The plan of the United States government to bring this result about by artificial means — paying the farmers to plow under cotton, not to plant wheat, and to kill little pigs — is now well known. The drought of 1934 however was probably even more effective in reducing the supply of agricultural products and raising their prices.

Natural events which tend to decrease the supply of wheat and so to increase its price are illustrated in the following quotations from an article by James A. Patten, himself the father of one of the greatest wheat corners in the history of the Chicago Board of Trade:

“In the Argentine, they commence to gather the wheat crop in November, and prior to that time I received word of a heavy frost down there. It was no secret.”

“About the end of August of 1908, Canada had a heavy frost. There were conflicting reports concerning that situation, too, but time showed that there had been a great deal of damage to the grain standing in the fields in the country to the north of us.”¹

“Though reports of black rust had seemed more important to the grain trade than the assassination of the Archduke Ferdinand,

¹ J. A. Patten, “In the Wheat Pit,” *Saturday Evening Post*, Sept. 3, 1927, p. 3.

when that tragedy occurred realization came with the successive declarations of war that occurred in August.

"In December the European demand sent the price up to \$1.28 $\frac{7}{8}$. During these months the fluctuations had been caused by the same facts that became first page headlines in the newspapers: the news that Turkey would join Germany and Austria sent the price up because it was realized this meant shutting off the Russian supply of wheat by the closing of the Dardanelles . . . the capture of a Turkish fort would send it down again; the sinking of two or three Allied warships would send it up; and so it went. The situation in the Dardanelles was always, throughout the war, a subject of great interest in the Chicago grain market."¹

The retail egg market. The reader may find it instructive to compare the history of the hypothetical egg market which has served us for illustrative purposes with the records of actual egg markets. The egg market is decidedly seasonal, although the placing of eggs in storage and the withdrawal of eggs from storage in periods of flush and slack production respectively have tended to even out the supply during the year.

The interrelation of these factors of production and price are clearly illustrated by the record on the next page.

The following figures relate to a number of markets combined, but they are indicative of what happens in any one market. Receipts of eggs rise during the first of the year until they reach a high point in April or May. Under the steady influence of increasing supply, egg prices drop off in the spring, reaching a low point at about the same time that receipts of eggs are greatest. Part of the excess production of the spring and early summer is withdrawn from the market for cold storage. Cold storage holdings reach a minimum about the end of February. They accumulate rapidly during April, May, and June, when prices are low, and reach a peak about the end of July. As receipts begin to fall off during the late summer, the price of eggs gradually rises, and storage holdings gradually decline. Prices rise during this period and continue to do so until the end of November. During the winter months the price

¹ J. A. Patten, "In the Wheat Pit," *Saturday Evening Post*, Nov. 5, 1927, p. 29.

EGG MARKET STATISTICS ¹

Year and month	Receipts five markets	Cold storage stocks end of month		Average retail price 51 cities
	(THOUSANDS OF CASES)	CASE (THOUSANDS OF CASES) ₁	FROZEN (THOUSANDS OF POUNDS)	CENTS PER DOZEN
<i>1934</i>				
January	808	50	49,910	30.1
February	1,165	90	39,181	26.9
March	1,824	1,208	38,679	24.7
April	2,051	4,640	62,632	23.8
May	1,927	7,819	93,947	23.4
June	1,452	8,965	116,058	24.2
July	1,009	8,961	121,564	26.5
August	828	7,938	111,994	31.6
September	665	6,803	99,951	34.8
October	655	4,633	88,715	36.7
November	588	2,380	76,073	39.7
December	642	648	64,879	38.5
<i>1935</i>				
January	750	39	52,726	37.7
February	858	34	39,413	36.7
March	1,488	1,508	39,516	28.6
April	1,866	3,901	59,313	29.6
May	1,963	6,366	84,741	—

of eggs gradually declines but stays on a relatively high level as compared with the summer months.²

The effect of perfect inelasticity of demand or supply. The reader will not have failed to observe that the conclusions reached as to the effects of changes of demand and supply upon price and quantity exchanged were in direct consequence of the respective slopes of the demand and supply curves. Before drawing a general conclusion we must take account therefore of peculiar conditions which give rise to vertical demand and supply curves.

Suppose buyers demand a fixed quantity of a given good for

¹ Receipts and cold storage stocks from *Survey of Current Business*, Jan., 1935, p. 42, and July, 1935, p. 44; prices are monthly averages of biweekly quotations in various issues of *Monthly Labor Review* (U. S. Bureau of Labor Statistics).

² For another record of the actual egg market, corresponding very closely with our theoretical example, cf. H. B. Vanderbleu, *Economic Principles, A Case Book*, 1927, pp. 110-116.

whatever price it can be obtained, as shown by the demand curve in Figure 27. Without following through the whole analysis in detail, it should be clear from a glance at the diagram that the

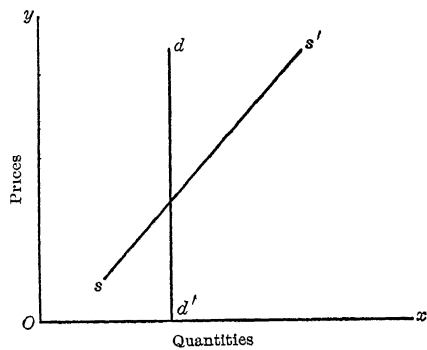


FIG. 27

vertical position of the demand curve makes no difference in the conclusions which we have previously reached, save for the one exception that changes in supply alone can obviously not cause any change in the quantity exchanged.

Similarly we conclude that, when the quantity offered is the same irre-

spective of price (see Figure 28), the general conclusions which we have reached are still valid except that changes in demand alone can cause no change in the quantity exchanged.

In view of our discussion of elasticity of demand and supply the reader will recognize in Figure 27 an example of a perfectly inelastic demand; and in Figure 28, an example of a perfectly inelastic supply. Here the quantity exchanged is determined by demand alone, or by supply alone, as the case may be; but the price in both cases is determined by the combined action of demand and supply.

The effect of perfect elasticity of demand or supply. To round out our study of the interaction of supply and demand we must also consider exceptional conditions represented by curves of the horizontal type exhibited in Figures 29 and 30. Suppose buyers — or

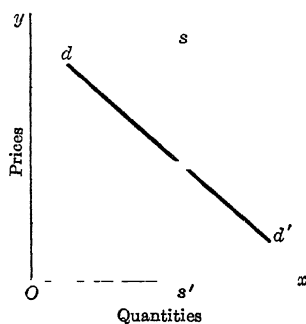


FIG. 28

more normally a sole buyer — to be able and willing to buy any amount at a fixed price, as for example, when a local canning factory offers to purchase at a stated price all the wild berries that are brought in. Demand in this case is represented by a curve such as

dd' in Figure 29. Another possibility is that an industry will have a horizontal supply curve, which means that sellers are able and willing to offer for sale any amount at a fixed price, as illustrated by the supply curve ss' in Figure 30. Such a supply curve, if interpreted literally, means that nothing will be offered at any lower price. In fact however as we have seen in our discussion of supply, an industry will have a horizontal supply curve only when an increase in output will be provided by the entry of new firms or the reentry

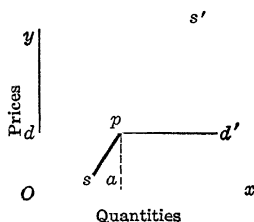


FIG. 29

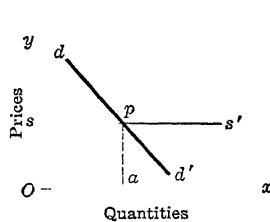


FIG. 30

of old firms into the industry, because in any perfectly competitive industry, the supply curves of individual firms must be upward sloping. Therefore a horizontal supply curve really refers only to the long period. In the short run a decrease in price will not mean a complete cessation of production and sales, because a number of existing firms will still be able to cover their variable costs and will continue in production, unless the reduction in price has been drastic.

In these two limiting cases of perfectly elastic demand or supply price is not determined by demand and supply together, but in one case by demand alone, and in the other by supply alone. In both these cases however the quantity exchanged is determined by the interaction of demand and supply.

The law of price change. Taking due note of the extreme cases described in the last two sections, we do not depart from the conclusion, previously reached, that when there is pure competition of sellers and a condition of increasing cost both the price and the quantity exchanged are in general determined by the joint action of demand and supply. Other conclusions thus far separately arrived at may now be summarized as follows: *When there is pure*

competition of sellers and a condition of increasing costs, (1) the direct effect of a change in magnitude of demand tends to be a like change in price; (2) the direct effect of a change in the magnitude of supply tends to be an opposite change in price; and (3) the direct effect of a change in the magnitude of either demand or supply tends to be a like change in the quantity exchanged. Indirectly, as we shall see, a change in the magnitude of demand may contribute to an opposite change in price.

When demand and supply both change. The foregoing generalizations were stated as tendencies, meaning that the results indicated will follow the respective causes unless some other cause interferes; that is, if other things are equal. In the illustrations thus far used we have assumed that other things were equal and have tested changes of demand or of supply, the other factor remaining unchanged. But it is quite possible and normal that changes in both factors occur at the same time. The hens may lay more eggs, thus increasing the supply, at the same time that the approach of hot weather causes a decline in demand. Or both demand and supply may either increase or decrease simultaneously. In such cases the effect is the resultant of the two influences, operating in accordance with the general law as stated above.

Thus let us suppose (1) that the demand increases while the supply decreases. Since the tendency of either change alone is to increase price, the combined effect will be an increase in price greater than would have resulted from either change alone. As regards the quantity exchanged, the two causes tend to neutralize each other. Whether the quantity increases or decreases will depend upon the relative magnitude of the changes in demand and supply. They may be such as exactly to neutralize each other, leaving the quantity exchanged the same as before. The various possible reactions to these changes in demand and supply are illustrated graphically in Figure 31, where dd and ss represent respectively the original demand and supply, while $d'd'$ and $s's'$ show the situation at the later date.

Similar demonstration of the other possible combinations is probably unnecessary. The conclusions may be stated as follows: (2) a decrease in demand and an increase in supply, tending each

to lower the price, would in combination cause a decline in price greater than would have been accomplished by either acting alone. On the other hand, these forces tend to neutralize each other as regards the quantity exchanged, which according to circumstances may be increased or decreased or remain unchanged. (3) Simultaneous increases in demand and supply tend to neutralize each other as regards their effect on price, which may rise or fall or remain unchanged. Since either of the causes now under consideration would, if acting alone, tend to an increase in the quantity exchanged, their combined effect must be an increase, and a greater

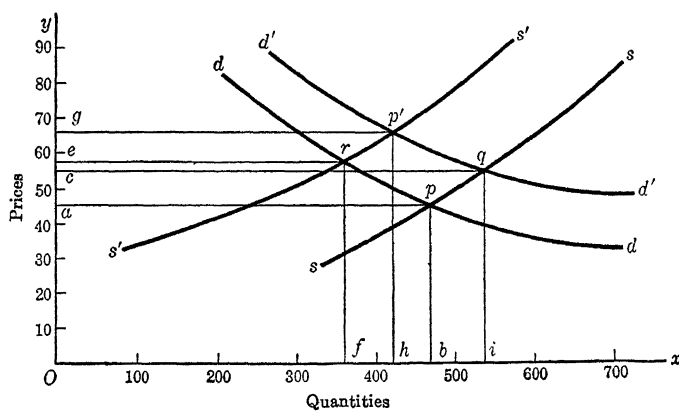


FIG. 31

increase than would have been caused by either alone. (4) Simultaneous decreases in demand and supply tend, as regards price, to neutralize each other, the resulting price being higher or lower than the original price, according to whether the predominant change is in supply or in demand, or being the same as the original price when the two changes exactly offset each other. On the other hand, since either of these causes alone would tend to a fall in the quantity exchanged, the two together must cause a fall, and a greater fall than would have resulted from either change alone.

The reader who has followed the discussion to this point will experience no difficulty in adapting the foregoing conclusions to the several cases of absolute elasticity or inelasticity of demand or supply.

Practical illustrations. The operation of these forces in various combinations may be observed at all times in the world of business and everyday life. Economic history furnishes striking examples of the tendency of like changes in demand and supply to counteract each other in their effects on price. Before the industrial revolution, cotton and woollen textiles were produced by primitive methods and at great cost. Producers, governed by costs, could offer only a small and limited supply. Since then the effect of invention in the textile industry has revolutionized the methods of production through the use of labor saving machinery, with consequent extraordinary reduction in costs. The result is a supply of such goods that makes the early supply appear utterly trivial. Prices have been greatly reduced, in spite of the increased demand of growing population, and the quantities exchanged have increased as the goods have come within the reach of the masses of people. Here the increased supply has been the dominant influence.

The following are more recent examples :

"The production of refined copper increased from 122,723 short tons in Jan. 1928 to 154,472 tons in Jan. 1929. Production reached a peak of 163,561 tons in March 1929. That demand was more than sufficient to absorb this increased production is shown by the decline in stocks of refined copper from 96,476 short tons in Jan. 1928 to 52,968. The price of copper ingot electrolytic in New York rose steadily from 13.85 cents a pound in Jan. 1928 to 21.26 cents per pound in Jan. 1929."

"Since April copper consumption has declined. In spite of a decline in production from 161,285 tons in April to 148,648 tons in August, copper stocks increased from 57,494 tons to 104,372 tons over the same period. The price of copper ingots, electrolytic, has sagged to 17.78 cents per pound."¹

"The production of crude petroleum in the United States increased from a monthly average of 39,349,000 in 1921 to 75,030,000 barrels per month in 1928. Imports declined from 10,447,000 barrels per month in 1921 to 6,632,000 in 1928. This rapidly increasing potential supply outstripped the increase in consumption as in-

¹ United States Bureau of Foreign and Domestic Commerce, *Survey of Current Business*, August, 1929, p. 44.

licated by run to stills of from 36,947,000 barrels per month in 1921 to 76,061,000 barrels per month in 1928. Stocks east of California rose from 179,888,000 barrels in 1921 to 368,026,000 in 1928 and the price at wells in Kansas and Oklahoma dropped from \$1.704 a barrel in 1921 to \$1.203 a barrel in 1928.”¹

Even in agriculture, where the importance of a scarce factor, land, would at first sight seem to offer little hope of increased product without increase in price, we find the historical record by no means so discouraging. Without changes in the conditions of supply, increase in demand with increased population would by now have led to ruinous prices of agricultural products. As a matter of fact there have been also epoch-making changes in the conditions of supply. The history of agriculture discloses important discoveries of new machines and new processes. The three-field system of cultivation employed on the medieval manor was not the most efficient. Later discoveries in the scientific rotation of crops and the scientific use of fertilizers have made possible an increased product at a lower cost. Modern labor-saving agricultural machinery has had the same effect. Thus invention from time to time pushes the supply curve to the right, tending to counteract the constant increase of demand.

The most perfect examples of changes in demand and supply which neutralize each other arise in connection with variations in the value of money. This subject will be investigated in a later chapter. The reader is now merely reminded that he will later find, in the study of money and prices, additional material having an important bearing on price changes, acting always however through the media of demand and supply and cost of production.

Decreasing costs. Thus far in this chapter we have studied price determination (always on the assumption of pure competition) under conditions of increasing and constant costs. These conditions permit the use of long run supply curves either upward sloping or horizontal. There is, as we have learned, no such thing as a downward sloping supply curve. Nevertheless, as was developed in Chapter XI, there are such things as decreasing cost industries, in which the cost of production for all the firms in the industry will

¹ *Ibid.*, p. 51.

be lowered as output expands and new firms enter. The reader will recollect also that it is possible for us to draw a down sloping curve which shows for any given output the price at which the industry will be in equilibrium. Although this is not a supply curve, it is a convenient way of representing certain significant characteristics of such an industry. For lack of a better name, we may refer to such a curve as a "quasi-supply curve."

Let us suppose that we are studying an industry which is operating under conditions of decreasing costs, as represented by the curve cc' , in Figure 32. The demand curve is dd' . In order to discover

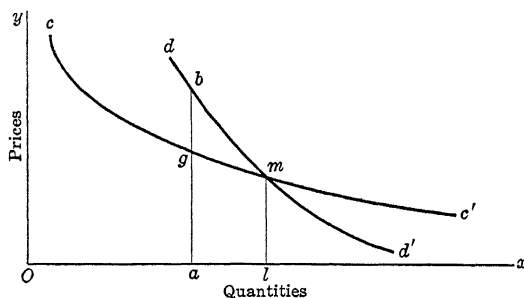


FIG. 32

how price will be determined in this situation, we may assume that initially the industry is producing an output Oa . What is the significance of the two points, g and b on the quasi-supply curve and the demand curve respectively? First, turning our attention to the demand curve, we know that the output Oa will command a price of ab . Since we have assumed that the industry is initially producing just this output and selling it in a competitive market, we know that initially the price will be ab . Turning now to the quasi-supply curve, we find that it gives us a second item of information. From the way in which the curve has been constructed, we know that the industry would be in equilibrium producing the output Oa if the price were ag . That is to say, if the price were ag the existing firms in the industry would have no incentive to alter their outputs, nor would new firms have an incentive to enter the industry, nor existing firms to leave it. We may infer all of this from our diagram, because the quasi-supply curve has been constructed to embody just such information.

Since the industry would be in equilibrium producing the output Oa at one price, and since the actual price is higher, we may infer that the industry, when we first turn our attention to it, is not in a position of equilibrium. Furthermore since the actual price exceeds that which would make the present output a stable one, we may expect output to expand. The expansion will come about, as always, in two ways. The high price will induce a larger output from existing firms than would the price ag , and the large profits that can be earned with the price ab will induce new firms to enter the field.

As the expansion of output under this stimulus proceeds we may infer again, by reference to our diagram, that two results will follow. As is always the case, the expanding output will force price downward. In the second place, because this is a decreasing cost industry, the cost curves of all the individual producers will be shifted downward, perhaps because raw material or some other element of cost is becoming cheaper, perhaps because the organization of the industry is becoming more efficient. Not only will these two developments occur, but we see from our diagram that price will fall more rapidly than the costs of the various firms in the industry. When output is expanded to Ol , the initial situation will have changed. The price will then have been forced down to lm , and reference to the industry's quasi-supply curve makes it clear that, at this price and with the output Ol , the industry is in equilibrium. Existing firms have no motive to change their output; new firms have no motive to enter the industry.

If we had picked as our starting point an output considerably in excess of Ol , the reverse reasoning would have been indicated. We would then have found that the market price was well below that required to keep the industry in equilibrium. The significance of this relationship between actual price and hypothetical equilibrium price is that firms would be induced to contract their output or to leave the industry. As output decreased the cost curves of all the remaining firms would be shifted upward. But simultaneously price would rise under the influence of contracting sales. Again the equilibrium position would be approached, and the move would come to an end with output at Ol and price at lm .

The reader must be reminded that the quasi-supply curve refers only to the long run. In the short run even a decreasing cost industry will have a true upward sloping supply curve, because in the short run there will not be time for new firms to enter the industry in considerable numbers or for old ones to leave it or even for existing firms to expand plant or alter the scale of their operations. Thus in the short run (as we have had occasion to note at several points) the supply curve of the industry is based on the upward sloping supply curve of the firms which compose it. For this reason the reactions discussed in the preceding paragraphs should be understood to refer to developments that take a considerable period of time. If in the above illustration price is initially well above that which would be required to keep the industry in equilibrium, this merely means that profits are being made which will sooner or later induce a substantial increase in output. The reaction would require a considerable period of time; business men would first have to become convinced that the profitable situation was likely to endure, and even then it would require some time before new firms could be organized and new plant and equipment built.

The reader will not have failed to observe that the reasoning in this case is precisely similar to that employed in previous sections. For many purposes of analysis the quasi-supply curve can be used in much the same fashion as a true supply curve. Nevertheless, we must emphasize again the fundamentally distinctive character of these two curves. The supply curve (upward sloping or at the limit horizontal) shows the relation between price (the independent variable) and quantity that sellers will offer (the dependent variable). The quasi-supply curve shows, on the contrary, the relation between quantity produced (the independent variable) and equilibrium price (the dependent variable).

Appraisal of foregoing methods of investigation: Theory and practice. Our investigation of the problem of competitive price has constantly required the employment of hypothetical assumptions and deductive reasoning, and our conclusions may appear to wear a forbidding aspect of abstraction. There are those who are distrustful of such generalizations. It may not be amiss at this

point to pause for a somewhat critical examination of our conclusions as to competitive price and the route by which we arrived thereat.

A scientific law is a statement of the necessary relation between things, or a statement of the effect which tends always to follow a certain cause. Every scientific law is thus a statement of a tendency, a statement that a certain result would always follow if certain conditions were fulfilled. If the conditions are not precisely fulfilled, the result does not precisely follow. In the demonstration and statement of a scientific law the conditions may be assumed with a precision and minuteness which can never be obtained in actual practice. There is thus always a certain discrepancy between the result as stated in the law and the way things work in practice. This is true of all sciences which deal with practical affairs. The science of physics contains a body of laws whose precise statements never agree exactly with the way things work for the civil engineer who has a railroad bridge to build. The practical chemist must always expect a certain difference between the actual results he gets and the precise statements of his formulæ. This does not mean that the scientific laws are false or inapplicable. All scientific law is hypothetical — if certain conditions exist, certain results will follow. The conditions are precisely stated. But it is never possible to include in such statement all of the conditions that will be found operating in the practical case. Every scientific law is, in some degree at least, an abstraction. Some of the more exact sciences, such as physics, chemistry, and astronomy, are able to go far toward the ideal of a complete statement of all the conditions. Yet there is always something in the practical result that could not be exactly foreseen and provided for in the formula.

These truths apply to all sciences, including economics; the various sciences differ only in degree. But it is true that economics labors under special difficulties in this respect, and we cannot safely neglect the task of examining our laws of competitive price from this point of view.

Market prices and imperfections of the market. It is evident that the laws of price which we have developed cannot explain every individual price in a competitive market, because we have so

far assumed not only that competition is pure but that the market functions with smoothness and efficiency as an institution. A "perfect market" in this institutional sense would be one in which all persons interested in either buying or selling the good were in uninterrupted communication with each other and were possessed of infinite knowledge of all the factors affecting demand and supply. In that sort of market the true market price would be instantly established and never departed from, and the laws of price would admit no margin of error. There would be none of the irregular scattering of prices which occurs in any real market before it settles down to equilibrium, and there would be none of the isolated irregular prices that may occur at any time in the real world. But of course no such market exists, although the closeness of the approach in some of the organized markets for dealing in corporation securities and certain staple commodities would surprise the uninitiated. Our laws relate to the equilibrium, or prevailing, or "market" price; they do not deny that goods are exchanged from time to time at other prices, for whose variations there may be insufficient explanation.

The student must also remember that for any good there will generally be, not one market and one market price, but many markets and many market prices, each one the result of forces of the same general character but of different magnitudes in the several markets. The prices of a given good in different markets will seldom be wholly unrelated to each other, as has been already suggested, but nevertheless they are distinct prices.

Economic friction. We must also note the presence of what is called "economic friction," by which is meant that complex of unknown factors which it is not possible to include in the statement of economic laws. These factors prevent perfect correspondence between scientific generalization and the practical event. The action of the laws of price is not smooth and instantaneous. Prices do not constantly oscillate back and forth like the sensitive magnetic needle on a storm-tossed ship. A considerable time may elapse before a change in demand or in supply works out its effect in price and quantity sold. Small changes may have no effect. Since conditions are constantly changing, perfect equilibrium is seldom reached

and never long continued. Even the tendency toward equilibrium is not smooth, regular, and constant, like the swinging of the pendulum back and forth across the vertical position or the oscillation of the magnetic needle. The economic adjustment of price and quantity exchanged is more or less irregular; it moves jerkily by fits and starts.

An analogy may be found in the operation of any machine. Everyone knows that the speed of an automobile, other things being equal, varies with the quantity of gas delivered to the motor. A smooth curve might be drawn showing this relation for any particular car. Yet in actual operation the correspondence between gas and speed will depart considerably from the precise continuous relation indicated by the curve. It will be impossible to detect changes in speed resulting from addition or subtraction of a minute fraction of a drop of gasoline. An appreciable amount must be added or subtracted to get an appreciable effect. And even then the effect will not correspond exactly to the quantity of gas. There is play in the motor bearings and in the transmission gears. The traction of the tires on the road is not perfect and constant. The response of the car to the throttle will not be absolutely smooth and constant; it will be slightly jerky and irregular.

The laws of price on the stock exchange. Possibly the closest approach in the world to a perfect market is the New York Stock Exchange. The buying and selling of securities in this market furnish excellent illustration of all the laws of price. They also illustrate the lack of perfect adjustment of cause and effect, even in the most nearly perfect market. Soon after the opening of the New York Stock Exchange at ten o'clock on the morning of August 22, 1935, 2,000 shares of the common stock of the United States Steel Corporation were sold at a price of $45\frac{1}{2}$. Of course the number of shares actually bought was equal to the number sold. But we cannot be sure that there were not some sellers willing to sell at this price but unable to find buyers. There may have been just one buyer calling for 2,000 shares and offering $45\frac{1}{2}$. If two or three brokers offered to sell him the 2,000 shares, he would pick one, presumably the one who called out his offer first. But wouldn't the others at once offer to sell at a lower price? Not necessarily. They

may not be willing to take a lower price, preferring to wait in anticipation of another buyer or not to sell at all. Conversely one seller may have offered the 2,000 shares at $45\frac{1}{2}$ and found several buyers willing to take them at that price but unwilling to offer more. Furthermore dealing on the New York Stock Exchange is normally in 100-share lots, and this obviously prevents a smooth and continuous adjustment of quantity exchanged. So we have the price determined, as well as the quantity exchanged, but there may not be a perfect "clearing of the market."

A little later this morning the price of United States Steel stock dropped to $45\frac{3}{8}$, with a sale of 200 shares. After three more sales at that price, it dropped to $45\frac{1}{4}$, and 800 shares were sold at that price in seven sales. The next sale was back at $45\frac{3}{8}$. So during the day the price oscillated back and forth between a high point of 46 and a low point of $45\frac{1}{8}$; the last sale of the day was at $45\frac{3}{8}$. During the day 25,200 shares were sold (the reports take account only of lots of 100 shares or more). The price changed sixty times during the day, ten times in the hour between 2 P.M. and the close of trading for the day. These are all facts, as reported on the stock ticker. They show a high degree of adjustment between demand and supply and price and quantity exchanged, yet not an absolutely perfect adjustment.

Shall we discard our theory? There are those who lose all faith in the general principles or laws of economics when confronted with facts such as these. We hear often enough: "What is the good of all your theories if they don't agree with the facts?" or "That is all right in theory, but it won't work in practice." The careful student of any science will not fall into such hasty conclusions. He will realize the true relation between a scientific generalization and the observed facts. He will see that there is always a lack of perfect correspondence, a certain "margin of error." But he will not therefore throw his general principles overboard. He will simply seek to determine as closely as possible the margin of error and to make allowance for it. Thus having learned how price would be determined under ideal conditions, he is ready to discover the discrepancies in practice, to explain them, and perhaps even to measure and allow for them. To throw overboard our economic

generalizations and their graphical illustration, because they do not fit perfectly the actual facts, would be as irrational as for the automobile engineer to discard for the same reason his curves of the relation between quantity of gas and speed.

Above all the careful student will not make the foolish statement that some principle "is all right in theory but won't work in practice." Any generalization which "won't work in practice" is not sound theory. But lack of perfect correspondence between theory and practical facts does not mean that the theory will not work. In spite of all allowance for discrepancies, it cannot be doubted for a moment that demand and supply did determine the prices at which United States Steel stock was sold on the New York Stock Exchange on August 22, 1935. There can be no doubt that, in spite of lack of perfect adjustment, the quantity of a good offered for sale will generally vary directly with the price, and that the quantity which buyers will choose to take does, by and large, vary inversely with the price. Changes in demand and supply certainly do have the effects on price which are laid down in our laws. These generalizations are true, making due allowance for the margin of error. They are also practically useful. Hardheaded business men, the first to recognize their lack of perfect application, have always relied, consciously or unconsciously, on the economic laws of price, and the up-to-date business man is going to greater lengths than ever before in the application of economic theories to the practical conduct of his business.

The students of the physical sciences and the practical men who work in those fields have generally outgrown the crude notions about "theory and practice" which still make trouble in economics and business. One of the most important tasks before the student of economics is to learn the real meaning of economic laws, their proper limitations, and the way to apply them to practical problems, realizing that only by precision in his concepts and rigid accuracy in his reasoning can he achieve success as a student of economics or gain the mastery of the economic forces which so largely dominate the world in which we live.

Competition. Competition is not popular with the ordinary business man. Business men would gladly eliminate or reduce

competition as it affects them. The protective tariff is advocated in order to remove the competition of foreign producers. The business man is no more friendly to the competition of his fellow citizens in other states or towns of his own country. The device of the protective tariff is not available, but he seeks the same end by trying to persuade his community to favor him as against the outsider. "Buy at home," "Patronize home industry," "Put your money in circulation where it will help your own community" are examples of the slogans used in this sort of campaign. The trolley cars carry signs reading "Patronize your neighborhood grocer." Back of all such persuasion is the desire to counteract the natural tendency of people to buy in the best market, which is the foundation of competition. It represents the struggle of producers to escape the controlling hand of competition. It is for the same purpose that producers come together and make agreements to eliminate competition between themselves. Thus there arise combinations of various sorts. And finally there may come monopoly, in which case competition between producers has ceased.

In an earlier chapter ¹ attention was called to competition as one of the foundation stones of the modern economic organization. We are now in position better to understand how competition works in the determination of price and as the regulator of industry. When competition has opportunity to work, it sees to it that no producer can long remain in business unless he is able to turn out goods at prices which neither exceed the relative marginal utility of the goods to the consumer nor fall much below the prevailing cost of production in the industry. The producer who has missed the public taste is ultimately eliminated, as is also the one who is not efficient enough to keep his costs as low as his competitors. The consumer, on the other hand, cannot have things to enjoy unless he is willing to pay for them prices normally equal to their cost of production, and he will not normally have to pay much more than that.

Competition thus assures to the consumers prices tending in the long run to be equal to cost. This assurance becomes progressively less as competition becomes less perfect. No such assurance is present under combination and monopoly. The causes affecting

¹ See Chapter VI.

price under conditions of imperfect competition and monopoly will be studied in the next chapter, and in a still later chapter we shall have to inquire into the special problems which arise when consumers are thus deprived of the protecting influence of competition.

EXERCISES

1. Assume the following demand and supply schedule for fresh eggs, medium size, in a certain city on a certain day :

<i>Price</i>	<i>Quantity that would be</i>	
	<i>Purchased</i>	<i>Offered for sale</i>
\$0.30	1300 dozen	200 dozen
.31	1140 dozen	370 dozen
.32	1000 dozen	530 dozen
.33	880 dozen	670 dozen
.34	780 dozen	780 dozen
.35	680 dozen	900 dozen
.36	600 dozen	1000 dozen
.37	530 dozen	1080 dozen
.38	475 dozen	1160 dozen
.39	410 dozen	1230 dozen
.40	350 dozen	1290 dozen

- (a) Plot the corresponding demand and supply curves.
 - (b) Draw, on the same diagram, a new demand curve indicating an increased demand and a new supply curve indicating a decreased supply.
 - (c) What is the new price of eggs? Explain the change.
 - (d) What is now the quantity exchanged? Explain the change.
2. (a) Plot again the demand and supply curves corresponding to the schedules of Exercise 1.
- (b) Draw, on the same diagram, new curves indicating an increased demand and an increased supply.
- (c) What is the new price of eggs? Explain the exchange.
- (d) What is now the quantity exchanged?
3. Using the definitions of demand and supply as adopted in this book, which of the following expressions is not permissible? Explain your answer and draw simple diagrams to illustrate in each case.
- (a) A cold winter increases the demand for coal.
 - (b) An increase in price would reduce the demand for coal.
 - (c) The supply of automobiles exceeds the demand.
 - (d) Favorable weather conditions increase the supply of wheat.
 - (e) Demand and supply are equal at the market price.
 - (f) The greater the demand, the higher the price.
 - (g) If manufacturers would reduce the prices of electrical appliances, demand would increase.

- (h) An armament boom will increase the demand for copper.
 - (i) An increase in demand causes an increase in supply.
 - (j) The higher the price, the less the demand.
4. What changes in demand alone or in supply alone or in demand and supply simultaneously would lead to the results indicated below? Draw simple diagrams to illustrate in each case.
- (a) An increase in price
 - (b) A decline in price
 - (c) An increase in quantity exchanged
 - (d) A decrease in quantity exchanged
5. What simultaneous changes in demand and supply would fail to cause (a) any change in price? (b) any change in quantity exchanged?
6. (a) Although the demand for farm wagons has fallen off enormously in the last thirty years, a wagon costs somewhat more today than it did in 1910. What is the explanation? Draw a diagram to indicate what has happened.
- (b) "Prices of automobile tires have steadily been reduced, in spite of an enormous increase in demand. This is a case of the failure of the law of demand and supply to function." Do you agree? Explain. Draw a diagram to illustrate what has occurred.
7. How would you explain the following price comparisons? Draw simple diagrams to illustrate in each case.
- (a) The price of cement was the same in 1930 as in 1900.
 - (b) The price of a standard grade of lumber was higher in 1930 than in 1900.
 - (c) The price of gasoline was much lower in 1938 than in 1908.
 - (d) The price of an ice refrigerator was the same in 1938 as in 1900.
 - (e) The price of sugar in 1938 was much lower than in 1860.
8. A manufacturer asks for bids to furnish him 100 machines of a particular type. He is prepared to pay any price up to \$150 apiece.
- (a) Draw a diagram illustrating this situation. What price does he have to pay?
 - (b) Draw supply and demand curves to indicate a condition under which there could be no sale.

XIV

PRICE UNDER IMPERFECT COMPETITION

Introduction. When we come to examine a market which is imperfectly competitive, it does not seem at first sight as though conditions of supply and demand exerted any very obvious control over the actions of individual sellers. What we see in a large department store or in the market for automobiles or in any other market where products are differentiated and sellers are few is that the sellers themselves set prices. In the first instance, at any rate, the suit which you buy in a store costs fifty dollars because the owner of the store chooses to place that price upon it. In other words, imperfectly competitive markets are usually those in which prices are named by the sellers; although this is not always the case, and the laws which govern price in such a market are exactly the same whether or not the seller sets the prices. However this should not blind us to the fact that, even in an imperfectly competitive market, prices are largely determined by forces beyond the control of individual business men.

Our study of the interaction of the forces of demand and supply under pure competition was much simplified because we could represent the behavior of sellers by a supply curve, just as we could represent the behavior of buyers by a demand curve. Putting these two curves together, we could then show how the actions of all the individuals composing the market determine price. We know that the individual producer under pure competition takes the price as given and governs his own actions in accordance with it. Thus the forces of demand and supply govern the individual seller under these circumstances through the agency of price alone.

Under imperfect competition, on the contrary, where the demand for each individual product is by no means perfectly elastic, the situation is very different. Market forces affect the individual producer by determining the conditions of demand for his own product instead of the ruling price. In graphical terms, they

establish the slope and position of the down sloping demand curve facing the individual seller instead of the position of the horizontal demand curve which represents the demand for his product in a situation of pure competition.

We have already studied the way in which the individual producer, selling in an imperfectly competitive market, will adapt his policy to the conditions of demand for his product as represented in an individual demand curve. In investigating the problem of market price in an imperfectly competitive market therefore we shall be studying the various ways in which market forces determine the conditions of demand for the product of an individual producer. This study will be the counterpart to our discussion in the last two chapters of the way in which the forces of demand and supply determine the ruling market price, at which each individual producer must sell his entire output or as much as he chooses to sell.

The rôle of substitutes : Their significance. Since we are now investigating the forces which determine the conditions of demand for the product of an individual producer, we must for a moment turn back and reconsider the nature of demand. In Chapter VIII, where we were discussing demand from the point of view of the consumer, we noted that his demand for any one commodity would depend not only upon his tastes and income but also upon the prices of other commodities and the quantities of them he already possessed. A demand schedule or curve always shows the relation between the price of a product and the quantity of that product which people are willing to buy.

But a demand curve does not imply for a moment that the quantity of a product people are willing to buy depends exclusively upon the price of that one product. Indeed the reader will recollect that it is impossible to draw up a demand schedule for any one commodity except upon the assumption that the prices of other commodities are given; any change in the prices of other commodities or in the incomes of buyers may be expected to alter the demand schedule of the product in question. From a demand schedule it is possible to discover at any given time how much more of the product people would buy if its price were reduced,

say ten per cent, but only if it can be assumed that other prices remain the same. This is important, since if other prices change at the same time, the demand curve will not tell us what will be the result of the price reduction.

These characteristics of a demand schedule or curve have special importance when we are discussing the particular products of individual sellers. Even in discussing the demand for low priced cars in general or meat or house-room, we were careful to note the dependence of demand upon the prices of other commodities; but when we turn to consider the consumers' demand for particular products — for Ford cars, or the clothes sold by one particular tailor, or house-room in one particular apartment building — it is essential to take more explicit account of the effect upon a buyer's demand of the availability of substitutes. It is quite sufficient to say of the demand for low priced automobiles in general that it will depend, not only upon consumers' tastes and incomes, but upon the cost of living, the prices of other kinds of transportation, and the like; but if we are concerned with the specific demand for Ford cars, we must note the very direct dependence of this demand upon the prices and qualities of Chevrolets and Plymouths, since these are both close substitutes for the Ford.

Of course the importance of substitutes for a given seller's product is just as great or even greater in a market with pure competition, but owing to the fact that under pure competition the output of one producer is not merely a close but a virtually perfect substitute for that of another, we paradoxically do not usually think of the one as a substitute for the other. Yet it should be clear that the reason that the demand curve for the grade A wheat produced by an individual farmer in North Dakota is perfectly horizontal is, not that the total demand for wheat is infinitely elastic, but rather that the grade A wheat produced by all the other farmers in the United States and elsewhere in the world is a perfect substitute for the grade A wheat that he produces. Since a perfect substitute for his product is available to all other buyers in the market, they will not pay a cent more than the market price for his output; and since his wheat is a perfect substitute for any one's else, he need not accept less.

Thus we conclude that the conditions of demand for his own product which govern the individual seller in an imperfect market are determined by the availability of substitutes, together of course with consumers' tastes and incomes. In this chapter we shall be mainly concerned with the way in which the actions of one seller affect others; that is, the relations among producers all selling products which are substitutes for one another.

Nature of substitutes. We must realize to begin with that there are substitutes for the products of all producers and that one commodity may be a substitute for another even though it is physically dissimilar and perhaps therefore not a very close substitute. In the broadest sense almost all commodities are substitutes for one another, since they are all competing with one another for the consumer's income. If he decides to spend more on his house, he will have less to spend on clothes and a car. Obviously however it is the closer substitutes that most directly affect the conditions of demand for any single product. These close substitutes may be other brands or makes of a similar product, as in the example of the automobile industry which we have frequently cited, or they may be physically dissimilar commodities which nevertheless render a similar service, such as gas and coal. What is significant for an analysis of price is the closeness of the substitute in this economic sense.

Influence of substitutes on the size of demand. The existence of substitutes affects the conditions of demand for a particular product in two ways: (1) by determining the size of demand and (2) by determining its elasticity. As regards the first, the number of substitutes and their prices will determine the general volume of demand for any specific product; that is to say, they will influence the position of the demand curve for any one seller's product. Let us consider the example of the competition between several grocery stores in a neighborhood. Each store will have certain regular customers attached to it by preference or habit, and it will also have an advantage over its competitors so far as the people who live in its immediate neighborhood are concerned. For the people within one block of grocery Store A but five blocks away from grocery Store B, Store A will be the more convenient. Because of these consumer preferences, the stores are selling in an imperfect

market; that is to say, Store A could charge slightly higher prices than Store B without losing all its customers and could cut its charges slightly below those of Store B without depriving the latter of all its customers. Each store will have a down sloping demand curve for any one of its products, rather than a horizontal one. In a sense, even though the actual physical goods which the several stores sell may be the same, they are to the buyers economically different goods because they are not perfect substitutes for one another.

We may most easily analyze the effects upon the demand for the products of Store A of changes in the availability of substitutes by inquiring what will happen if Store B sharply cuts the price of one of its products, let us say coffee. Clearly B's action will have the effect of attracting a number of A's customers. Many people who live closer to A will now be induced to buy their coffee from B because of his lower price, even though it will be slightly less convenient for them to do so. The demand curve for A's coffee will be shifted downward and to the left; that is to say, A will sell less at any given level of prices after B's action than he would have before. He will be compelled either to reduce his own price to meet that of B or else to sell less coffee.

The appearance of a new competitor in the neighborhood will affect Store A in the same way as if B had cut prices. Suppose a new store, C, is established only three blocks away from Store A. Store C will now be the most convenient for a considerable group of people who formerly traded with either A or B, and in time Store C will build up a clientele of customers who regularly deal with it. Clearly the positions of the demand curves for A's coffee and B's coffee respectively will be altered. Both will be shifted downward and to the left. We may generalize and say that, from the viewpoint of any seller, a reduction in the price of available substitutes for any given product or an increase in the number of substitutes will lower the demand for the product.

Influence of substitutes on the elasticity of demand. Substitutes will influence, not only the general volume of demand for the product of a particular seller, but also the characteristics of demand, notably its elasticity. The way in which this influence is exercised

may be made clear with the aid of another example. Let us take our previous illustration but change the conditions by assuming that before Store C appeared in the neighborhood A and B were selling quite dissimilar goods. Perhaps A had been selling goods of lower quality and catering to people who could not afford to spend too much for food, whereas B had been catering mainly to wealthier patrons. Under these circumstances, it would require quite a sharp cut in prices on B's part to attract many of A's customers. The demand for each store's goods would be inelastic. Now let us suppose that when Store C is established it sells goods of practically the same quality as those sold by A, that is to say, Store C also caters to the trade of the poorer members of the community. Store A now has, not only more competition than before, but also the competition of products and services which are closer substitutes for its own in that they are economically more similar. In this new situation it may still be true that a considerable price cut on B's part would be required to attract many of A's or C's customers, but it will emphatically not be the case that this relation holds as between A and C. Since these two stores appeal to similar sorts of customers, and since the products sold in one are economically very similar to those sold in the other, a small price cut in one store will attract many customers from its rival. Conversely a small increase in prices in Store A will now drive away a larger proportion of its remaining customers than it would have when B was the only other grocery. In other words, the appearance of Store C has not only reduced the demand for A's products but has at the same time made that demand more elastic.

Conclusion. We may sum up these considerations in two propositions concerning the demand for the products of an individual producer selling in an imperfectly competitive market as follows: (1) *A reduction in the price of available substitutes or an increase in their number will diminish the demand for the product of a single producer and vice versa.* (2) *An increase in the similarity or "closeness" of the nearest substitutes will increase the elasticity of demand for the product of an individual producer, and vice versa.* Armed with these two propositions, we may now proceed to study the process of competition and price formation in an imperfect market.

Interaction of sellers' policies. The reader will recall from his perusal of Chapter IX the way in which an individual seller in an imperfect market would behave when faced with a given and known demand for his own product. The next step in a study of the process of price formation in an imperfect market is to investigate the way in which the activities of many independent sellers react upon one another and govern the behavior of each. The foregoing discussion of the significance of substitutes suggests a most important observation; namely, that under imperfect competition the individual seller cannot take the demand for his product as given; *i.e.*, as independent of his own price-setting acts. In fact the business man must always take account of a factor which does not appear in a demand curve; namely, the probable behavior of competitors. Especially is this so if the demand for his own product is highly elastic, so that a small cut in his price would (other things being equal) largely expand his sales. For then that same action would be bound to do considerable injury to his competitors and provoke retaliation from them. Such retaliation, in the form of price cuts by competitors, would reduce the demand for his own product. It is common knowledge that the most difficult problems which a business man must solve in determining his price and sales policy concern the changes in the demand for his product to which his own actions may indirectly give rise by causing retaliatory action on the part of competitors. In general any one producer feels the impact of the activities of his competitors, because their policies determine in considerable measure the demand for his own product.

For this reason we must distinguish clearly between two kinds of information or assumption, upon the basis of which business men's decisions are made. The first is information about the demand for his product, dealing specifically with the ease with which consumers will be induced by small changes in price to spend more money on one product and less on others. The demand for Fords is undoubtedly very elastic, because the three very low priced cars are generally believed to be similar in quality, and it is a reasonable presumption that if the price of Fords were cut ten or fifteen per cent, while those of Chevrolets and Plymouths remained the same, many more Fords would be bought.

The second body of information does not directly concern what has been defined as the demand for the individual producer's product, but rather the probable reactions of competitors to changes in the individual's own policy. As we shall see, competitors' reactions may be expected to differ widely from one industry to another, and the importance of this second body of information will depend upon the particular industry under discussion. But it is impossible to analyze the functioning of an imperfectly competitive market unless we take account of the fact that sellers base their decisions, not only upon estimates of the demands for their products, but also on some knowledge of their competitors' behavior; in other words, upon estimates of the indirect effects of their own actions.

Price determination: Long run and short run. In this as in the previous chapter we distinguish two problems, to be taken up in turn. The first or short run problem relates to the actual process of competition and price formation. In the short run it is true in an imperfectly competitive industry, as in one that sells in a perfect market, that the number of firms and the scale of their operations may be treated as fixed. The second problem concerns the influences that operate in any group of producers in the long run and recognizes the fact that the number of firms in the industry may change or the whole scale of operations of existing enterprises may be altered through the building of new or the retirement of old equipment.

Price determination in the short run: Price cutting. It will simplify the problem if we make use of an example similar to that used above in our discussion of substitutes. Let us assume again that we are considering two firms which produce close substitutes and sell them in the same market, such as the two grocery stores selling coffee of the previous example. We have already examined the way in which the demand for A's coffee will be affected by the price of B's. Suppose that when we first turn our attention to these two firms, A has succeeded in adopting a price which just equates marginal cost with marginal revenue. He has therefore no incentive either to raise or to lower his price. But let us suppose that B has not succeeded in adopting so successful a price policy;

he finally arrives at the conclusion that if he were to cut his price five per cent he would expand his sales substantially and increase his revenue more than his costs. In other words, he discovers that his marginal revenue is still substantially in excess of his marginal cost. Presumably when he has come to this decision he will lower his price, let us say five per cent, in the hope of thereby increasing his profits.

Whatever the wisdom of the price cut from B's point of view, it immediately has certain repercussions upon A. As B reduces the price of his product, he will begin to cut into A's trade, and the demand curve for A's product will shift downward and to the left in accordance with one of the propositions stated above. A now finds himself out of adjustment. So long as the price of his product remains unchanged, the decline in demand will sharply reduce his sales. With this smaller output A's marginal costs will be lower, while his marginal revenue is likely to be the same or higher. So it is very likely to be profitable for him to meet at least some part of the decline in demand with a price reduction. Especially will this be true if the elasticity of demand for each producer's product is high, since in that case B's price reduction will attract a large number of A's customers.¹ When A cuts his price, he will in turn have upset B's original calculations, because after A's price cut B's demand curve will have been shifted downward. By precisely the same reasoning we conclude that B may then have an incentive to cut once again. For if his initial price cut would have enabled him just to equate marginal revenue and marginal cost, and if A's action has now reduced the demand for B's product, B will presumably find marginal revenue still in excess of marginal cost. This process of competitive price cutting is familiar enough to business men, and once started it may continue for a considerable time.

Limitations: First limiting factor. If we study this problem in further detail, we can distinguish three limiting factors which will

¹ Of course it is conceivable that as the demand for A's product declines it will also become less elastic, so that A will have no incentive to cut his price and will find it most profitable simply to yield up a number of customers to B. But it is much more probable that, after B's price cut and the decline in A's sales, A's marginal revenue will not have declined so much as his marginal costs.

ultimately halt price cutting and with reference to which we can extend our conception of equilibrium in the imperfectly competitive industry. The first factor will be operative even if each seller disregards entirely the second of the two kinds of market information which we have distinguished. So let us see first what would happen if A and B ignored the fact that each price cut provokes retaliation. It is clear that although each price cut by one firm reduces the demand for the other firm's product, and by the same token its sales, nevertheless the first firm's sales are so increased that the total sales of the two firms taken together will be increasing as these two firms together cut into the demand for the more remote substitutes for their own products. Now in a realistic situation price cutting by any two firms producing close substitutes is likely to spread beyond these two, and there is likely to be a certain amount of price cutting in the whole field in which their products are sold. Nevertheless if the products of firms A and B are very much alike, and if there are no other firms which produce close substitutes, the price cutting will expand the sales of both A and B, at the expense perhaps of many of the other things which consumers buy. For instance if these two firms were the only ones selling electric ice boxes in the United States, competitive price cutting between them would expand the sales of electric ice boxes at the expense of furniture, automobiles, radios, etc. Since the sales of both firms will be expanding as the process of price cutting goes on, their marginal costs will rise, and each will be approaching more and more closely a position where marginal revenue is equal to marginal cost. The process of price cutting will be limited as the firms approach a condition where neither of them has any incentive, as firm B did initially, to carry the matter further. The final situation is one in which there is no possibility of further price cutting, because neither firm, even disregarding the actions of its competitor, has any incentive to lower its price. Therefore so long as each seller ignores the indirect effects of his own price cut, this is a stable equilibrium position.

Second limiting factor. However it is in fact most unlikely, as we have seen, that sellers will completely disregard the retaliation of competitors, and if they do not, there will be some incentive for

one of them at least to put his price higher than it would be when each seller has equated marginal revenue and marginal cost. The reasoning may be made clear if we suppose that price cutting has gone to the bitter end, so that each firm has equated marginal revenue and marginal cost. It will then pay any one firm to raise its price slightly, for the following reason. If A raises his price, his action will increase the demand for B's product. But B has already carried production to the point where marginal cost is equal to marginal revenue, so that an upward shift of the demand curve for his product will almost certainly induce him to raise his price. This means that A's initial increase in price will have induced B to follow suit.

Here the indirect effect of A's own action partially offsets the direct effect, just as it does when price cutting is in progress. A's raising of his price, after price cutting has gone to the limit, induces B to follow. This raises the demand for A's product, so that his original action, because of its indirect effect, will deprive him of less business than his demand curve would have led him to expect. Even though his marginal revenue and marginal cost were equal before he raised his price, the move will probably turn out to have been profitable; and he will take this action if he gives consideration to the second kind of market information. We conclude that a situation in which each firm has equated marginal cost and marginal revenue is not a stable one, unless the various firms in the industry pay little attention to the effects of their actions upon their competitors. Our analysis would thus lead us to expect that competitive price cutting would generally not be carried to the bitter end or, if it were, that prices would subsequently be raised by producers to some point above this lower limit.

It may very well be that when a stable situation has been reached, marginal cost will be equal to marginal revenue for some of the competitors and below it for others. In the above example, for instance, we have assumed that A was the first one to raise his price and that B subsequently followed suit because the demand for his product has increased. Thus we would expect to find that in the end B had equated marginal cost and marginal revenue and A's marginal revenue exceeded his marginal cost. In general we

can say that prices would be reasonably stable at some level above that which would enable all producers to equate their marginal costs and marginal revenues.

Third limiting factor. The reader should recollect that we arrived at this conclusion by assuming that producers take cognizance of the indirect effects of their actions, but that they are nevertheless real competitors. We have taken it for granted that A and B in our illustration do not enter into any sort of agreement with one another. This is however by no means always true. Realistically therefore there is a third factor which will set a limit to competitive price cutting; though one that cannot and need not be described in such detail.

In an industry where competitors are few and where the marginal cost curve throughout a large part of its length lies well below the average cost curve, there is constant danger of what is known as "cutthroat competition," under which the lower limit to price cutting may not be reached although all of the firms in the industry are suffering losses. This will be especially true if the demand for the output of the industry as a whole; that is, for the whole series of substitutes produced by the various firms involved, is highly inelastic. Under these circumstances almost inevitably the producers will through tacit agreement or understanding strive to avoid price cutting. Each realizes that, if price cutting once starts, the process may be difficult to stop. Although marginal revenue may be well in excess of marginal cost for everyone in the industry, there will thus be a strongly developed "public opinion" (perhaps fostered by a trade association) to prevent the initiation of a price war.

Though an uneasy truce may be maintained in this fashion it is by no means accurate to call the position one of equilibrium, and indeed it is bound to be highly unstable. For until the limit, at which every firm has equated marginal revenue with marginal cost, has been approached at least a few of the competitors in the industry would profit by a price cut and an expansion of output if only they could be sure that other firms would not follow suit. A change in the policy of one firm or the impact of a depression is all that is needed to start the price cutting process. The absolute

limit to this process will not be reached, as we have seen, until the first of the three limiting factors described above becomes operative; that is, until each firm has so expanded its production that marginal revenue is equal to marginal cost. Price cutting will usually however stop short of this point, or if it has been carried this far, the process may be reversed and prices set somewhat higher.

Conclusion. Where competitors are comparatively few and where they sell individualized products each one of which is in some degree a substitute for the others, their price behavior will tend to have its well defined characteristics. (1) *Prices change infrequently, because of the general dislike and fear of price cutting.* (2) *Prices tend to be stabilized somewhat above the lower limit to which price cutting could carry them; i.e., the point at which prices have been cut until every firm has equated marginal revenue with marginal cost.* (3) *There will be an inherent instability in the price structure of an industry unless prices have been stabilized near this lower limit.*

The relative importance of these conclusions will depend upon the type of imperfectly competitive industry under consideration. In the cement and steel industries, where competitors are few but the products sold by the different producers are very close or even perfect substitutes for one another, the second type of market information will have very great importance. A very small cut in price by one cement company virtually compels the others to follow. Therefore no intelligent executive would cut the price of cement without taking account of the inevitable reaction of his competitors. Conversely in such fields of business as the retail trade competition is imperfect mainly by reason of the individuality of products and the element of consumer preference for a particular seller, and not by reason of the fewness of competitors. Even though the number of sellers is large, the demand for an individual seller's product may be so inelastic that he can wisely disregard the reactions of competitors when considering price changes. For instance a price cut by one grocery store in a neighborhood may not have a great effect on the demand of the corresponding product of any one other grocery, because the store which has cut its price

will draw only a few customers from each of the other stores in the district. Here a moderate price cut might provoke little or no retaliation. The business man in this type of industry is not likely to base his actions to so great an extent upon estimates of what his competitors will do.

Price determination in the long run. Having studied the way in which market forces control price determination in the short run, we must take account of the fact that in the long run the number of firms and the amount of plant and equipment in an industry are generally not fixed.

Our previous conclusion that prices would not be stable if they were far above the level which would equate marginal cost and

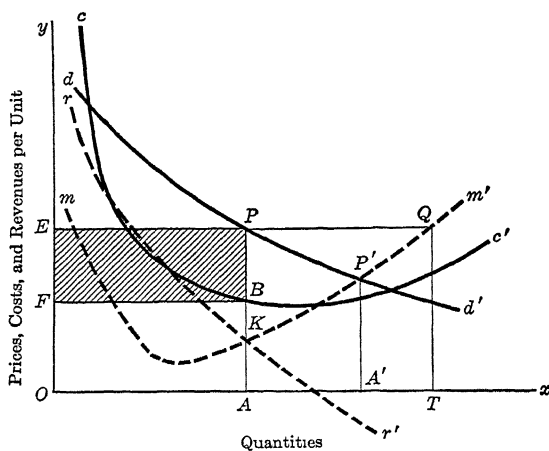


FIG. 33

marginal revenue does not mean that established firms will not be able to earn very substantial profits in the short run. We may illustrate the point by referring back to the individual firm represented in Figure 17 on page 240, which is here reproduced as Figure 33. In that diagram the firm has set a price AP , such that it can sell the output OA and equate marginal cost and marginal revenue precisely. Although price is not above the level which would be the lower limit to price cutting, it turns out that the firm is earning very substantial profits, represented by the area $EFBP$. If this firm is doing so well, and if others in the industry are

similarly fortunate, the mere fact that it is selling in an imperfect market does not mean that it is free from outside competition. Unless the industry of which it is a part is highly organized to prevent the entrance of new competitors, or unless the profits that the firm is earning are abnormal and not expected to endure, new competitors will be attracted into the same line of business by the high level of profits. Since we are studying an imperfect market, it is highly probable that our producer's product is individualized and protected by brand name or patent or trade mark, so that he will not face the competition of other producers turning out products which will be perfect substitutes for his own. Nevertheless new competitors will appear and will offer to the public products which presumably will be close substitutes for that of the producer in question. As new competitors enter the field, it is likely that some of them will offer closer substitutes than have hitherto been available. It is probable therefore, in accordance with the two propositions which we derived from our study of the significance of substitutes, that the demand curve dd' will shift downward and to the left and will at the same time become more elastic. As this process continues the profits of our firm will diminish. The firm will always strive to maximize its profits, but its best efforts will be unable to prevent this decline.

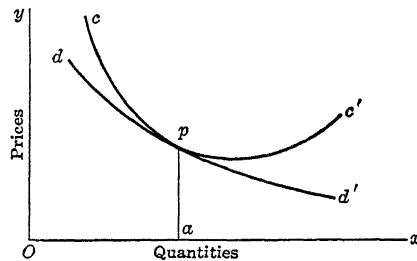


FIG. 34

A position may finally be reached which can be illustrated by Figure 34. In this situation, the demand curve dd' has shifted so far to the left that, instead of cutting the average cost curve at two points and lying above the cost curve between those two points, it will be tangent to the cost curve at one point, p . This means that if the firm produces and sells the output Oa , which will command the price ap , it can just cover its average costs and avoid losses. But since for any other output the cost curve is above the demand curve, any other output will involve loss. This shows that the

output Oa is the one which will maximize the profits of the firm, and we know that if we were to draw in the marginal cost and marginal revenue curves they would intersect at some point on the line ap . By the time the firm we are studying has reached this position, it is probable that the general level of profits earned by firms producing substitutes will also have been forced down to a very low level. There will therefore be no incentive for new firms to put still other substitutes on the market. In this particular sector an equilibrium will have been reached, exactly similar to the equilibrium achieved in an industry under pure competition when the entrance of new competitors has driven the market price down to the point where it just covers the average costs of the various firms in the industry.

This discussion shows that even where competition is imperfect, in the sense that there is a down sloping instead of a horizontal demand curve for the product of each individual producer, the established producers in a field will not be safeguarded from competition. They will, it is true, have much greater freedom to determine at any particular time what prices to charge. But if competition is at all keen, and if they earn substantial profits, the profits will attract new competitors into the industry. The appearance of new firms will lower the demand for the products of those already in the field, force them to push their sales more intensively, and ultimately make it impossible for them to continue to earn abnormal profits, just as surely as if they were selling in a market where pure competition prevailed.

Examples of this limitation upon the freedom of the producer in an imperfectly competitive market may easily be found. There are only three established makes of low priced automobiles, and there is very considerable individuality to each. Therefore as we have several times had occasion to note, the automobile industry provides an excellent example of operation under conditions of imperfect competition. Yet no one would deny that competition among the three companies is very keen. Except in boom years when output is enormous they do not earn an abnormally high rate of profit. Furthermore in the periphery of the industry there are a number of marginal concerns which have been through

bankruptcy or are close to it. If the "Big Three" were able to make very substantial profits year in and year out, it would be possible for one of these smaller competitors to put out as good a car at a lower price and thus to move the demand curves for the three major makes downward and to the left so that their profits would in time be reduced.

Conclusions. Our study of price determination in an imperfectly competitive market must inevitably seem inconclusive and unsatisfactory compared with the elegant analysis of supply and demand which preceded it. In a market characterized by pure competition buyers and sellers trade in a single standardized commodity. Therefore the study of price determination under conditions of pure competition is a study of the way in which the price is determined by market forces. When we turn to the study of imperfect competition, there is no such simple relationship. We are able to show that each individual producer will set the price of his product according to the conditions of demand which exist in the market. But although we can study the way in which market forces determine the conditions of demand which face the individual firm, that does not mean that every firm will charge something that could be called "the price" for "the product of the industry." For under imperfect competition there is no well defined "industry," and we are dealing, not with a single "product of the industry," but rather with a series of different products which are close substitutes for one another.

We will usually expect to find that in an imperfectly competitive market the price set by one producer will not be the same as the price set by another. This is in no way remarkable or difficult to explain, since as a rule there will also be differences in quality between the output of one producer and another. In the typical market of this sort there is intense competition. This competition tends to act on existing firms in such a way as to force them to set prices slightly above the point at which marginal cost and marginal revenue would be equal. If when producers have been forced into this position they are still earning substantial profits, we would then expect new producers to enter the field until these profits had been eliminated.

This conclusion is parallel to that which we reached in our study of markets with pure competition. We saw there that the pressure of competition, by making it impossible for firms to sell their output except at the going market price, would induce them to expand output to the point where marginal cost was equal to price and that, if in the long run the existing firms would make abnormal profits under these circumstances, new firms would enter the industry until the level of profits had been reduced to normal. The main difference between pure and imperfect competition relates to the mechanics of the market and to the function and behavior of price. In the market with pure competition there is a single ruling price; competition is always pure price competition. In an imperfectly competitive market there are many prices for different kinds of products; competition is by no means purely price competition, and a decline in demand does not always mean a lower price, nor an increase in demand a higher price. And — most important of all — price is usually substantially higher than it would be under conditions of pure competition.

EXERCISES

1. Assume the following conditions of demand and cost of production for a low-priced automobile:

<i>Price</i>	<i>Sales</i>	<i>Marginal revenue</i>	<i>Average total cost</i>	<i>Marginal cost</i>
\$875	100,000	\$875	\$1,310	\$310
780	200,000	685	813	315
700	300,000	540	650	325
635	400,000	440	573	340
584	500,000	380	530	360
540	600,000	320	508	400
500	700,000	260	500	450
465	800,000	220	503	520
435	900,000	195	516	620
410	1,000,000	185	542	780

- Plot on a sheet of graph paper the four curves of demand, marginal revenue, average total cost, and marginal cost.
- What is the best output for the manufacturer to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?
- What is the profit per unit and the total profit under these conditions?

2. A new manufacturer of automobiles enters the low-priced field. The additional competition reduces the demand for the product of the first manufacturer, described in Exercise 1. Assume that the new condition of demand for his product is as follows:

<i>Price</i>	<i>Sales</i>	<i>Marginal revenue</i>
\$785	100,000	\$785
700	200,000	615
630	300,000	490
572	400,000	398
524	500,000	332
485	600,000	290
450	700,000	240
418	800,000	194
390	900,000	166
365	1,000,000	140

- (a) Using the chart constructed in Exercise 1, plot the new demand curve and the new marginal revenue curve.
- (b) What is now the best output for this manufacturer to produce? Reading from the chart, at what price can this output be sold? What is the average total cost of producing this output?
- (c) What is the effect of the appearance of a new competitor on the price of the product of the first manufacturer and on the volume of his profits?

3. The Federal Trade Commission, a government agency charged with the suppression of monopoly and the preservation of competition, has devoted considerable effort to the prevention of "brand imitation." Does an increase in the similarity or "closeness" of available substitutes tend to make competition more or less perfect? Explain.

4. (a) The Harvard Stadium has a seating capacity of about 60,000. The Athletic Committee sets a price of \$2.00 per ticket for the game with Shady Rock and the stadium is half-filled. Draw a diagram to illustrate this situation. Reading from your diagram, at what price could the stadium have been sold out?
- (b) The Committee sets a price of \$5.00 for the Yale game and is unable to fill all orders for tickets. Draw a diagram to illustrate. Reading from your diagram, how many tickets could have been sold at \$5.00, if the seating capacity of the stadium had been large enough? What price could the Committee have charged and still have sold out the stadium?

.....

XV

INTERRELATIONSHIP OF PRICES. PRICE THEORY IN PRACTICE

Interrelationship of prices. We have been investigating the way in which prices are determined by the forces of demand on the buyer's side of the market and supply and cost of production on the seller's side, under various conditions of pure and imperfect competition. It has been recognized throughout this investigation that these immediate forces which directly determine price are themselves the resultants of a great variety of forces. Among these forces, acting thus indirectly upon the price of any good in any market, will always be found the prices of other goods or other markets.

In the last analysis all prices are interrelated in some degree. We have seen that any individual's demand for any good is a complex result, affected, among other things, by the prices of all other commodities and services which he considers or might consider buying. Under conditions of imperfect competition the demand for any seller's product is affected by the prices offered by competitors. The prices of the products of industry are related to the prices at which the services of the several factors of production — land, labor, and capital — are to be obtained. The limits of this book do not permit full investigation of the innumerable price relationships here suggested. We may however profitably give special attention to some of the more direct and fundamental ways in which prices are related to each other.

The present value of future income: Capital and income. The theory of price is incomplete without inclusion of the most fundamental of all price relationships; namely, that between the value of future income and the present value of capital goods. We have learned that all production requires capital.¹ To obtain income man

¹ In the present discussion we shall use the term capital in the broadest of the several meanings which have become irretrievably attached to it; *i.e.*, as a syno-

must have capital goods, and it is the production of income which alone gives significance to capital. In a primitive society this effort to secure capital takes the direct form of devising and constructing the weapons, tools, buildings, and other instruments by means of which the productivity of man's labor is multiplied and his enjoyable income increased. Under such simple conditions therefore it is customary for the same man both to bear the final costs of capital — the labor of making the instruments and of keeping them in repair — and to consume its benefits. Neither capital nor income is frequently bought and sold; hence neither acquires an exchange value, and there is no tendency to dissociate capital from income and regard it as an end in itself. Capital is viewed in its proper light as a means toward an end, and, like other means, each unit of capital is more or less prized according to its efficiency in achieving the end in view; that is, its power to produce income for its owner.

In the social order of today there are to be found many instances of this direct and visible relation between capital and income. Men sometimes gather materials and with their own labor construct some simple capital instrument whose benefits will accrue to themselves directly. The farmer may cut and shape a handle for his axe or build a fence or wagon shed. A blacksmith may make some of his own tools. In such cases the relation of capital to income is clearly grasped; one sees that it is the expected income which bestows on capital its importance and that the owner's appraisal of the capital is but a reflection of his evaluation of the benefit which he hopes to obtain from it.

Property and capital. But these simple cases are not typical of present-day society. So complex has our economic structure become that we often lose sight of the fact that capital has no significance when divorced from the benefits which it produces. Chief among the attributes of our social system which have this effect is the fact that the ownership of capital has been subdivided into a large variety of property rights, which are bought and sold in great

nym for wealth, thus including land as well as man-made instruments. The term is generally so used in discussions of investment and when it is desired to give prominence to the distinction between wealth, the fund, and income, the flow. Cf. Chapter I and the footnote on page 88 (Chapter V).

numbers. Those property rights are usually represented by documents of different kinds — stocks, bonds, mortgages, contracts, and the like — which frequently have stamped upon them some nominal valuation (for example the \$100 “par value” on a stock certificate), which misleads the owner into the belief that these documents have a “normal” or “true value” in themselves. The development of many agencies with the purpose of facilitating trading in these property rights — such as banks, brokerage and investment houses, stock exchanges, and so on — has aided in beclouding the true relation of capital to income.

When certificates of property are bought and sold in great numbers, they fall into the hands of people who have neither knowledge of the capital of which they are evidence of ownership nor interest in the direct income which that capital produces. The owner of stock in a railroad, for example, need not, and generally does not, know anything at all about the technique of the capital instruments which make up a railroad system, nor need he consciously share in the enjoyment of the benefit (transportation) whose production is the reason for the existence of those particular capital instruments. He may merely look at the par value of his stock and conclude that for some reason this bit of paper has a true value of that amount. He would be at a loss to explain why these concrete capital instruments are worth \$100, unless indeed he had recourse to the customary explanation that they probably cost that much to produce. In general it is true that the average man has very vague and incorrect notions of the way in which the value of capital and of property rights in capital is determined. It is the purpose of this chapter to show that the value of capital and property rights depends fundamentally upon the value of the income which the capital produces and to investigate the process by which this factor enters into the determination of the value of capital.

Property valuation. Let us first consider the relation of the value of property to the value of capital. As we have learned, property is but a right to the future benefits of wealth (capital). Property does not produce these benefits but merely determines who shall get them. The benefits are produced by the capital

which underlies the property; it is the function of the property right to divide this income among the owners of the capital. It is apparent therefore that the principal reason for the existence of the certificate of property is that it affords a convenient method of transferring ownership in capital from one man to another and of subdividing the ownership so that many may share in the income produced by a given set of capital instruments. These certificates are bought and sold as representatives of the valuable thing (capital) in which they transfer ownership; hence the action of the buyers and sellers in the market will result in giving the certificates the same value that in the opinion of these buyers and sellers should be placed on the capital which they represent. If for example, a manufacturing concern has issued one thousand shares of stock and these shares sell in the market for \$150 each, there can be but one reason for this; namely, that in the opinion of investors all the assets of the concern are worth \$150,000 over and above the amounts owed to the creditors. The fact that the stock has stamped upon it a nominal value of \$100 or \$200 or \$50 or any other figure will not determine its market value as long as the capital which the stock represents is believed to be worth \$150,000. The reader will recall that the subject of valuation of capital stock was investigated in more detail in Chapter V.

Present worth. The real problem therefore is to discover how the value of capital goods is determined. Here we must repeat that capital is desired only because of the income it produces and is bought only as a means of obtaining this income. Let us suppose that a given machine will last ten years and produce a net income of \$100 a year. If this machine is offered for sale to the owner of a factory, he will consider that by owning it he can increase his income during the next ten years by the amount of \$100 a year, or \$1,000 in all. Whatever price he is willing to pay for the machine will be a measure of the present value he puts upon this thousand dollars, distributed as it will be through the ten years to come. To change the illustration, we may suppose that an acre of land is offered for sale to a man who has reason to believe that it can be made to produce a net income worth \$10 a year forever. In this case too the price offered by the buyer for the land will be his esti-

mate of the *present worth* of an annual income of \$10 which continues indefinitely. Or again an investor may be offered a bond which will be redeemed for \$100 at the end of five years and in the meantime pay \$5.00 a year interest. Such a bond would be viewed as an opportunity to receive a series of money payments during the next five years: \$5.00 one year from now, another \$5.00 two years from now, another \$5.00 three years from now, \$5.00 more four years from now, and a final payment of \$105 after five years have elapsed. The amount of money that the investor is willing to part with today in order to buy the bond must be an indication of what he considers this series of future payments to be worth at the present moment. This is the essential thing in calculating the value of any capital good or of any property right in capital goods. It amounts to placing a present value on an expected future income of a certain size and form and duration.

Present goods more desirable than future goods. But by what process can we link present and future values together? It is a well-known fact that men do not consider a sum of money which will accrue in the future to be worth its face value now. The investor referred to in the preceding illustration will not calculate the present value of the bond by simply adding together the different payments which will accrue in the future to its owner. Such a calculation would give the bond a present value of \$125; and the buyer paying this price for it would eventually get back just the amount of money that he paid, though he would have to wait five years for most of it. To change the illustration, if one offered for sale a promissory note which would pay its owner \$100 twelve months from today, he would find no one willing to pay \$100 for it now. Nor would this reluctance to buy this future income at its full face value in present money be due only to risk of loss. If the \$100 were locked up so safely that all doubt regarding its security were removed, still no one would be willing to pay out \$100 now and be content to get it back after twelve months had passed. In the mind of the ordinary man the present value of the future dollar is less than that of today's dollar; hence sums of money which accrue in the future will be bought only if they can be had for less than their face value. And the price that will be

paid for capital which gives its owner the right to receive income in the future will be lower, by some substantial amount, than the value of this future income when it accrues.

Interest. If asked why he will not pay \$100 now for the right to receive \$100 one year hence, the practical man will probably answer that he is unwilling to lose interest on his money. Because of the existence of interest, it lies in the power of any man who has present wealth to exchange for future wealth to make the exchange on terms which will bring him a future return of larger value than his present payment, and as long as this opportunity to gain interest on the money they lend is open to all men, no one will consent to pay its full future value in present money when buying income. *Interest may be defined as the premium which present goods command over similar goods due in the future.* The man who lends \$100 for one year at five per cent interest exchanges \$100 worth of present goods for \$105 worth of next year's goods, the premium of \$5.00 ($\frac{5}{100}$ of the present sum) being required to give the future goods a present value equal to that of the present goods which the lender surrenders. As the interest rate rises and falls the premium which must be added to make these two amounts — one present, the other future — of equal present value rises and falls accordingly. It is therefore the *rate of interest* which determines how much larger the future return must be than the present sacrifice in order to cause the exchange of the two to occur.

The link between future and present values. We need not at this point attempt to explain why interest exists or how the rate of interest is determined; these are problems to be solved later in our study. For the present purpose it suffices to accept the existence of interest as a feature of the present business world. Our immediate concern is to show that the rate of interest supplies the link between future and present values which enables us to translate the one into the other.

Given the rate of interest, it is a simple matter to calculate the equivalent future value of any amount of present money. When the interest rate is 5%, for example, \$100 invested now will be worth \$105; *i.e.*, $\$100 \times 1.05$, one year from now; \$110.25; *i.e.*, $\$100 \times (1.05)^2$, two years from now; \$115.76; *i.e.*, $\$100 \times (1.05)^3$,

three years from now; and so on. This translation of present into future values through the medium of the interest rate may be reduced to a formula, as follows:

$$A_n = P(1 + r)^n$$

in which:

A represents the future value, any number of years hence

P represents the present value, or "principal"

r represents the rate of interest

n represents the number of years.¹

This same formula may be adapted to the reverse problem of determining the present worth of a given future value by merely rearranging its terms, as follows:

$$P = \frac{A_n}{(1 + r)^n}$$

the symbols having the same meaning as in the preceding formula. To illustrate, assuming again an interest rate of 5%, \$100 due one year hence is worth now \$95.23; *i.e.*, $\$100 \div 1.05$; if due two years hence, it is worth now \$90.70; *i.e.*, $\$100 \div (1.05)^2$; if due three years hence, it is worth now \$86.31; *i.e.*, $\$100 \div (1.05)^3$; and so on.²

¹ The demonstration of this formula is as follows:

At the end of the first year, the amount (A_1) is the sum of the present value, or principal, and one year's interest; that is,

$$A_1 = P + Pr = P(1 + r).$$

That is, the amount at the end of the first year is the principal multiplied by $(1 + r)$.

This amount, $P(1 + r)$, is the new principal for the second year. The amount at the end of the second year (A_2) will be found in exactly the same way as for the first year; that is, by multiplying the principal at the beginning of the year by the factor $(1 + r)$. Thus

$$\begin{aligned} A_2 &= P(1 + r)(1 + r) \\ &= P(1 + r)^2. \end{aligned}$$

In the same way, we find at the end of the third year the amount is $P(1 + r)^3$, at the end of the fourth year, $P(1 + r)^4$, and so on for any number of years.

² The examples in the text assume that the interest rate is so much *per annum* and that interest is "compounded" annually. The formula holds however for any unit of time. All that is required is that the rate of interest be stated with respect to the same period as is used in compounding. For example it is required to find the amount of a loan of \$5,000 for two years at the rate of 1% a month, interest compounded semi-annually. Since the interest is compounded every half year, we must take a half year as the unit of time. Then we must call the rate of interest, instead of 1% per month, 6% per half year. Instead of saying the loan is for two years, we must say it is for four half years. Then our formula becomes $A = 5,000 (1.06)^4$ and, when we solve, we find that $A = \$6,312.38^+$.

The formula which determines the present worth of a given future value will serve likewise to place a present worth upon a series of future values. What is the present worth of a series consisting of \$200 due one year hence, \$500 due a year later, and \$100 due at the end of the third year, if the rate of interest is 5%? By means of the formula, $P = \frac{A_n}{(1+r)^n}$, we find the present worth of these future amounts to be respectively \$190.48, \$453.52, and \$86.31. The total of these three amounts, \$730.31, is the present worth of the series of future values. By this process, one evaluates the ordinary bond. Thus a 5 per cent \$1,000 bond due in 7 years is valued as a series of 6 annual payments of \$50 and a final payment of \$1,050 at the end of the seventh year.

It is frequently important to evaluate a uniform perpetual series of future values, as in the case of a perpetual government bond. The general formula for the present worth of a uniform perpetual income is $P = \frac{i}{r}$, in which P stands for the present worth, i stands for the periodical payment, and r stands for the rate of interest.¹ Thus assuming an interest rate of 3%, the present worth of a 4 per cent \$1,000 perpetual bond (like the British consols) would be $\$40 \div \frac{3}{100} = \$1,333.33$.

The discount process. Although it is possible by means of the interest rate to calculate either future values from present or present values from future, and although in the business world there

¹ This formula is derived, as a special case, from the general interest formula, remembering that, in this case, the interest is not compounded but paid each year, so that the principal is brought back each year to the original amount, and the amount at the end of each year (just before the interest payment) is the sum of the original principal and one year's interest; *i.e.*, for each year $A = P + i$. According to the general formula,

$$P = \frac{A_n}{(1+r)^n}$$

Assuming a single interest period, $n = 1$. Then

$$P = \frac{A}{1+r}$$

But

$$A = P + i$$

Hence

$$P = \frac{P + i}{1 + r}$$

$$P + Pr = P + i$$

$$Pr = i$$

$$P = \frac{i}{r}$$

is continual use for both of these calculations, it is the latter process which has the greater importance. This is true for the reason stated at the beginning of this chapter; namely, that the economic activity of mankind is directed toward the obtaining of income and all income which has power to give value to capital lies in the future — not in the past. The valuation of capital instruments and property rights is incidental to this struggle for income, being a product of the incessant buying and selling of the means by which income is produced. Hence the calculation of the present values of future incomes of every conceivable variety as regards size and duration is one of the commonest events in the business world. The name given to this calculation of present values of future incomes is the *discount process*.

Summarizing the preceding discussion, we may say that this process involves these two fundamentals: (1) a knowledge of the probable future value of the income, (2) a knowledge of the interest rate. The first of these fundamentals is the basis of the calculation — the future value whose present worth it is desired to obtain. The second supplies the mechanism for determining the present worth; the process being to divide the future value by one plus the rate of interest raised to a power equal to the number of years which must elapse before the future income accrues.

Capital and income. It will be observed that the relation of capital to income may be looked at from either of two points of view; namely, that of the physical relationship and that of the value relationship. From the physical point of view it is capital which produces income, an orchard produces a certain number of barrels of apples; a factory, so many pairs of shoes or other units of output; a house, so many years' shelter, and so on.

But the value relationship is the opposite of this. The value of the capital does not produce the value of the income. The value of the apples gathered from an orchard will be established in the market in which the apples are sold after they have been produced. The price of shoes will also be determined in the market by the forces of demand and supply which operate when the shoes are sold. So it is with the product of any capital instrument; its value will be governed by the laws of market price as explained in

the preceding chapters. Setting a value upon the capital will not avail to dictate the terms upon which its product can be sold. A given set of capital instruments may be very productive, in the sense that it turns out a large stream of products, but if these products cannot be sold at profitable prices, the value of the capital instruments will be very low, or even non-existent. This is true of capital when the demand for its products has declined or disappeared. Abandoned farms, factories, and railroads bear testimony to the fact that the value of capital goods is derived from that of their income, just as worthless stocks and bonds are painful evidence that property rights have no value in themselves.

To summarize briefly: Capital in the physical sense produces income; income derives its value from the conditions governing the market at the time it is sold or enjoyed; the estimated amount of future value of income governs the value of the physical capital by which the income is produced. That there is also a causal relation proceeding in the opposite direction, from cost of production of capital goods to the value of income, will be made clear in a later section of this chapter.

Discount and the laws of price. It is essential that we understand exactly how this process of discount is related to the general laws of price which we have previously developed. We have found that under conditions of pure competition prices are in general determined by demand and supply and that prices under imperfect competition are determined by decisions of the sellers in conjunction with the conditions of demand and with consideration of actual and potential costs. We have learned that back of demand is marginal utility, and back of supply, cost of production. Now it appears that values are determined by discounting future expected incomes. Are we faced with two conflicting principles?

To make our question concrete, suppose we ask: Why does a particular tract of land sell for \$200? Is it because this is the price determined by the forces of demand and supply in the market, or is it because this is the present worth of an expected perpetual income, say \$10 a year discounted at 5 per cent? The answer is "Yes" to both questions, and the absence of any conflict will be apparent as soon as we inquire into the part which discounting

plays in the market. Values are not determined directly by discounting. In the simple numerical examples which we have thus far employed, we assumed certain future values and we assumed a certain rate of interest. In the practical world, neither of these essential factors is objectively determined. The future incomes — just because they are future — are uncertain, more or less. The rate of interest is the rate at which any given person will discount the future. Every potential buyer and every potential seller must come to his own decision as to what these factors are in any given case. A, on the basis of the information available to him and his best judgment, decides that the tract of land in question will probably bring forth such and such future incomes, that these incomes will have certain values to him, that he is prepared to discount the future at a certain rate of interest; with these factors thus estimated, he goes through the discounting process and determines how much he is willing to pay for this piece of land. B makes similar estimates, arriving possibly at different figures from A at every point, and finally emerging with the price he is willing to pay. So of all other possible purchasers. And the present owner of the land must himself go through a similar process in order to determine at what price he will be ready to sell. When the potential buyers and sellers meet in the market, the discounting process has already taken place in the mind of each. If the tract of land in our example sells for \$200, it is because the seller considered it worth something less than that to him, and the purchaser considered it worth something more. In arriving at his judgment, each employed the discounting process. The discounting process is itself one of the forces that determine demand and supply, which in turn determine price. Discount, so far from being in conflict, is itself included in the principle of demand and supply.

We recognize of course, as in all such theoretical analyses, that every person does not consciously go through the mental process here visualized. Some buying and selling, as we have seen, is more or less irrational. But back of demand and supply will always be found human judgment discounting expected future incomes. Even where — as appears to be the normal case on the speculative

stock and produce markets — dealers seem to make their decisions, not with any regard to discounted future incomes, but on the basis of expectations as to what prices will be in the future, the fundamental principle is still in operation, though somewhat obscured. For in the last analysis, back of all estimates of future prices, there is always someone's forecast of future income and its discount at a rate of interest. The dealers are simply cutting corners to arrive at the result. The market rate of interest itself is a composite judgment of the relation between future and present values.

Discounted income and cost of production. There are a great many goods that are staples in the market. They are constantly being reproduced as a matter of routine business practice; day after day they appear in a steady stream upon the markets of the world. The producers of these goods, after they have offered their supplies for sale, must be content with a price determined by the laws of market price, in the operation of which the discount process will govern the demand of the buyers. But what if the estimate thus placed by the buyers upon any given quantity of the goods differs from the cost of producing this quantity? The discrepancy will be resolved, as we have learned, through the power which the producers have either to avoid future costs by giving up or diminishing the production of these goods or, on the other hand, to increase future costs by increasing their output. Their decisions on this point will, after an interval of time, exert an influence upon the value of the capital goods they are producing. If they produce an increased quantity of the capital goods, the services rendered by them will increase in quantity and command a lower price; hence the discounted value of the capital goods will fall. On the other hand, if the producers reduce their output of the capital goods, the peculiar benefits which flow from these goods will become scarcer and their price will rise, hence raising the discounted value of the capital goods. In other words, it is the willingness or unwillingness of the producers to bear future costs which, by affecting the value of the income from capital, brings marginal utility and marginal cost into harmony.

In the case of standard and staple goods, the markets for which have settled to a point of relatively stable equilibrium, the pro-

ducers may confidently take the present price as an indication of what the future price will be. If the present price of the good stands very much above the cost of making it, existing producers will be tempted to expand their output; while at the same time other enterprisers will be tempted by the prospect of large profit to engage in its production. As the number of instruments in use increases, the value of the income which one instrument can earn will decrease, and consequently the discounted value of the instrument will fall. If the industry is subject to the condition of increasing costs the marginal cost of producing the instrument will rise as attempts are made to supply it in larger quantities. As a result the discounted value and the marginal cost, which at the outset were different from each other, will tend to draw together.

On the other hand, if the present price of the capital instrument had been below the marginal cost, enterprisers desirous of avoiding the losses caused by this situation would abandon the production of the instrument entirely or contract their output of it. This restriction in the supply of the instrument would cause the income it produces to grow scarcer and higher in price, with the result that the discounted value of the instrument would rise, and the diminution in the production of the capital instrument would reduce the marginal cost of production. Again there would be a tendency for the two — marginal cost and discounted value — to approximate each other.

In the case of capital goods which are not being currently reproduced, discounted value may part company from cost of production indefinitely. Then it is discounted future income, not cost of production, that determines the value. The construction of dwellings, office buildings, hotels, and factories will not be undertaken unless the builders expect the selling price to be as high as the cost of production. But after the costs involved in the erection of a given building have been borne and the building is offered for sale, the price will be determined by the buyers' estimate of the present worth of its future products. A summer resort constructed at great expense in an unpopular place will command a price lower than its cost of production because the discounted value of the net income to be derived from its operation is low. For the same reason

an office building erected in a decaying town will have a value below its cost of production.

Prices of the same good in different markets. Turning now to a different type of price relationship, let us suppose that the price of wheat in Chicago is \$0.80 a bushel and in New York \$1.05 and further that the entire cost of transporting a bushel of wheat from Chicago to New York is 15 cents. There is evidently a profit to be made by buying wheat in Chicago, shipping it to New York, and there selling it; and there is a class of business men ever on the alert to take advantage of such opportunities. Such dealers will immediately enter the market, seeking the profit of 10 cents a bushel which the existing conditions indicate. But these very operations will inevitably change the conditions. The dealers enter the Chicago market on the demand side, increasing the demand for wheat and so raising its price. In New York, on the other hand, they appear upon the seller's side and tend to increase the supply and lower the price. Even so, the margin of profit, though declining, continues for a time, and a considerable quantity of wheat will be transferred from Chicago to New York. The prices on the two markets will approach each other until finally the difference is only 15 cents; let us suppose the Chicago price has risen from \$0.80 to \$0.84 while the price in New York has fallen from \$1.05 to \$0.99. The two prices are now in equilibrium with respect to each other, differing by exactly the cost of transportation. In such manner the prices of the same good in different markets may react upon each other.

Such reaction is obviously possible only where there would otherwise be a difference in price greater than the cost of transportation from the low price market to the high price market. If the price of brick is \$10.00 per thousand in New York and \$8.00 per thousand in Chicago, while the cost of transportation between the two cities is \$20.00 per thousand, these prices are evidently independent of each other, since no one will be tempted to buy brick in one market for sale in the other.

The extent to which prices are affected by the prices of the same good in other markets thus depends upon the cost of production in relation to price. Why were the commodities which

entered into the long distance trade of the medieval merchants so generally articles of luxury — jewels, silks, fine wines, spices, perfumes, etc.? The answer is that heavy transportation costs made it unprofitable to bring to market from remote places anything that could not be sold at a high price. Cheap and bulky commodities, such as farm produce, could be sold profitably only in the markets close to the place of production. Since that time the transportation system has become so extraordinarily efficient and the price of transportation so low that the province of trade has been enormously extended. Cheap and bulky goods, such as wheat and bananas, are shipped thousands of miles over land and sea to be placed at the disposal of distant consumers. With each step in the progress of lower transportation costs, the realm of these price reactions is widened. For many products, such as wheat, cotton, the precious metals, etc., the markets of the whole world are thus tied together. This subject will be further investigated when we undertake the study of interregional trade, at which point we shall better realize how vitally these price relations affect the well-being of mankind.

Prices of competing goods. In the winter of 1925-26 there was a long continued strike of the miners in the anthracite coal mines of Pennsylvania, which for a considerable time virtually stopped the flow of such coal to the markets. The consequent decline of supply raised greatly the price of anthracite. As a consequence, people sought other kinds of fuel, such as bituminous coal, coke, etc. This meant an increased demand for such fuels and in consequence higher prices. Suppose there is in a certain market a sudden increase in the supply of beef. Beef prices will decline in consequence, and more people than usual will buy beef. This will entail a decrease in the demand for lamb, let us say, and the price of lamb will therefore go down. These events illustrate another sort of price relationship; *i.e.*, the relation between the prices of substitutes or competing goods, with which the reader has already become acquainted through earlier chapters. We have the general principle that the prices of competing or substitute goods are so related that they tend to rise and fall together.

Prices of complementary goods. Suppose some change in conditions of production should bring about a material increase in the price of bread. People would naturally incline to diminish their consumption of bread, seeking to find similar nutritive satisfaction from other kinds of food. The result might be, in accordance with the law of competing prices, a rise in the price of meat or vegetables. But what, if any, would be the effect upon the price of butter? Butter is chiefly used in connection with bread. Using less bread, people would want less butter. The demand for butter would decline and its price would fall. Similarly a fall in the price of bread, by causing its increased use, would increase the demand for butter and tend to raise its price. This is one example of a pair of complementary goods, meaning goods which are used only or chiefly in conjunction with each other. There are plenty of other examples. In the extreme case, the use of the one good is possible only in connection with the other; for example tennis rackets and tennis balls are completely complementary goods. It is evident that the prices of complementary goods tend to move inversely to each other, the reverse of the relation between the prices of competing goods. For example horses and mules are substitute or competing goods; their prices tend to rise or fall in unison. Horses and wagons are complementary goods; their prices tend to move inversely.

Prices of competing cost goods. The price relations of competing goods and complementary goods arise out of the reactions of the buyers; they are on the demand side of the market. There is another relation between competing goods (substitutes) which we have not as yet encountered. Labor is normally devoted to those pursuits which offer the greatest return. Consider the case of a small town in which there is one important manufacturing industry making, let us say, grindstones. A rise in the price of the product of this enterprise will cause its owner to seek to expand his output; for this purpose he will require more labor, which he will draw, by offer of higher wages, from the dairy farms and market gardens of the neighborhood. The resulting shortage of agricultural labor will cause a decreased supply of milk and vegetables in the local market, with a consequent rise in their prices. Thus a rise in the

price of grindstones has caused a rise in the price of vegetables. Products such as these, which both require the same cost element, make competing demands for the factors of production; competing in this example for the use of labor. They may be called "competing cost goods." Their prices evidently tend to move up and down together. This relation is quite distinct from the one which arises from the demand side; that is, from the fact that a rise in the price of one good increases the demand for all substitutes.

Prices of joint cost goods. Where cattle are slaughtered for beef, hides are made available, as well as horn, tallow, and a great number of other products. These products appear jointly as the results of the same production costs. Having determined to produce a certain quantity of one, the industry will inevitably bring forth the corresponding quantities of the other products. Where the products are of unequal importance, the less important is called the *by-product*. For example the growing of cotton necessarily yields a crop of cotton seed. The latter has its uses, but in general cotton would not be grown for the sake of the seed; that appears incidentally to the production of the primary crop. Such products as these are said to be the results of joint costs, and their prices have a definite relation to each other. Let us suppose that something causes the price of cotton to be especially high. To take advantage of this, the planters will increase their production of cotton, thereby inevitably increasing the supply of cotton seed. The price of the latter will consequently be reduced. A fall in the price of cotton would discourage production, thus leading to a decline in the supply of cotton seed and a rise in its price. The prices of joint cost products thus tend to move inversely, being the same relation as was found to exist between the prices of complementary goods. In recognition of this analogy, joint cost goods are sometimes called "complementary on the supply side."

Prices of "tandem" goods. Where one good is the material from which another good is made, there exists still another sort of price relation. If the price of wool rises, there is caused thereby an increase in the marginal cost of producing woollen cloth. This will cause a reduction in the supply of woollen cloth; *i.e.*, at any given price the quantity offered for sale will be less than before.

Hence the price of woollen cloth rises. This is a simple example of a principle that is a commonplace of everyday observation; *i.e.*, that the prices of manufactured goods tend to vary directly with the prices of the materials that enter into their production.

Price theory in practice: Practical aspects of the theory of value.

The very considerable time which we have devoted to study of the theory of price is justified by the importance of these concepts in every department of business and practical affairs. Inasmuch as the modern social system rests upon private ownership of wealth and free exchange, problems of value are at the heart of the practical enterprises of mankind. This appears nowhere more clearly than in the field of investment. The ownership of corporations (in the broadest sense) is expressed in their stocks and bonds. These securities are bought and sold, in some cases daily or even hourly on the exchanges, in other cases more or less rarely and privately. Values are determined thus by the action of demand and supply, and back of demand and supply is always the process of discount or capitalization of income. The buyers of any particular stock present a certain demand, because they have first separately made their estimates of the future incomes to be expected from this stock and secondly have capitalized these expected future incomes, at what they consider proper interest rates, in accordance with the discount process which has been investigated in this chapter. The supply of the particular stock, presented by the sellers, is determined in precisely the same way. Consciously or unconsciously, every buyer or seller or owner of corporate securities is making practical use of value theory in reaching his decisions to buy or to sell or to hold.

That the values of corporate shares sometimes appear to be quite out of harmony with the current dividends, so far from being denial of the validity of our theory of value, is the best proof of its truth. Before the sensational stock market crash in the United States in the autumn of 1929, the prices of American stocks generally were very high. For example the common stock of the United States Steel Corporation reached its high point of $261\frac{3}{4}$ on September 3. The annual dividend rate being 7 per cent, purchasers at this price and owners who held at this price were apparently satisfied

with a return of only $2\frac{2}{3}$ per cent. Now the fact is that American stockholders are not satisfied with any such return as that. The explanation is that these investors were not so naïve as to discount only present income; they were acting in accordance with sound theory and discounting the future incomes expected to flow from shares in the United States Steel Corporation, based upon surpluses already accumulated and expected future earnings. September, 1929, witnessed the culmination of a long period of buoyant optimism among investors, and there was a general anticipation of high future earnings for American corporations. The reaction is seen in high values and low apparent yields all along the line. To mention only two more cases, the common stock of the American Telephone and Telegraph Company sold on September 19 at $310\frac{1}{4}$, showing a yield just under 3 per cent on the basis of annual dividends of 9 per cent. General Electric Company common stock sold at 403 on August 20, although its dividend payments in 1929, including extra dividends, amounted to only \$6.00, indicating a present yield of less than $1\frac{1}{2}$ per cent. These are among the aristocracy of "gilt-edged" securities, not the really "speculative" stocks.

That these rosy views of the business future were exaggerated and not warranted by the facts was the opinion of some of the more conservative observers as early as the summer of 1929, and the correctness of this opinion was demonstrated by the collapse of values that occurred in October and November. United States Steel dropped from its high price of $261\frac{3}{4}$ (on September 3) to 150 on November 13; American Telephone and Telegraph came down from the high peak of $310\frac{1}{4}$ (on September 19) to $197\frac{1}{4}$ on November 13; and between August 20 and November 13 General Electric experienced a dizzy fall from its proud eminence of 403 to the modest price of $168\frac{1}{8}$. At the new price, Steel stood to yield on the basis of the past year's dividends $4\frac{2}{3}$ per cent, American Telephone and Telegraph, $4\frac{1}{2}$ per cent, and General Electric, $3\frac{1}{2}$ per cent.

The important lesson for our present purpose is that it was the investors' and speculators' estimates of future facts, not their theory of value, that was wrong. This is of course not to say that every investor or speculator goes through a careful process of

estimating and discounting future corporation earnings; many jump immediately to the conclusion that security prices will go up or down on the basis of past quotations or rumors. But back of all such evidence rests the judgment of those who are, to the best of their ability, predicting and discounting future incomes. And this brings out clearly the principle that all value is the resultant of human judgment, human estimates of future events. Obviously no one is sure of the future. In any broad market, such as the stock market, there are dealers of all degrees of knowledge and judgment, from those who by resort to all possible sources seek to inform themselves of conditions and exercise the best judgment that is humanly possible, to those who act, without knowledge or intelligence, on the basis of rumors or "tips" or mere guesswork. The resultant demand and supply are the consequences of the composite judgment of these buyers and sellers, weighted in proportion to their respective resources. Knowledge and ignorance, sound judgment and recklessness, conservative investment and pure gambling, all combine to determine prices, through the chain of cause and effect which we know as the laws of price or value.

The overworked term "value." Although the concept of value as used in the science of economics is thus basic in practical affairs and so recognized by practical men, there have developed in certain administrative fields other concepts, closely related to value though not precisely the same, which generally go by the name of "value." Thus the accountant finds it necessary to assign values to the various items in his statements. Starting with the basic notion of value, accounting practice has evolved certain concepts which serve its purposes but depart more or less from the economic notion of value. Thus items of wealth, such as land, buildings, machinery, etc., may be carried on the books on various bases of cost: cost less depreciation, replacement cost less depreciation, etc., even though these "values" may be known to be higher or lower than the considerations for which the articles could be sold. "Goods in process" of manufacture are "valued" on the cost accountant's books on such bases as cost of materials plus direct labor cost, with or without a share of overhead, even though

in certain cases there might be no possibility of sale of such partly finished products except to another manufacturer in the same line, in the absence of which these goods might have no real value except as scrap. In such cases as this the term value is frankly used to mean something different from the economic concept of value. It is always unfortunate when the same term is used with diverse meanings, but a clear understanding of the usages employed will generally suffice to avoid confusion.

The true value of a corporation may generally be best shown by the total market value of its securities. This may however differ widely from the net value of its assets as shown by the balance sheet. The discrepancy will be due to a variety of causes. In the first place, the accountants' valuations may depart from market value, as already seen; and, such as they are, they may not be kept strictly up to date. Secondly the balance sheet may not include certain property rights of an especially intangible or elusive nature, such as the character of management, certain franchise or contract rights, the "good will" of an established clientele, etc. — factors which are combined under the title of "going concern value" and upon which the future earnings of the enterprise may in large part depend. The balance sheet valuation, though serving effectively its own purpose, may thus depart materially from the economic concept of value. That the latter is the fundamental one is demonstrated by the very fact that buyers and sellers of the corporation's securities arrive at values different from those shown by the balance sheet.

In the practical task of regulating the rates of public utility corporations, such as the railroads, the telephone and telegraph companies, gas and electric companies, etc., it is found necessary to ascertain "values" of the capital invested in order that the legally determined rates may permit a "fair return" on the investment. In this task the various official regulatory bodies, such as the United States Interstate Commerce Commission, the state railroad and public utility commissions, and the courts, have developed a technique of "valuation," which scarcely pretends any longer to seek to arrive at the economic concept of what the property would sell for. What is actually sought is a "rate base" which shall

effectively serve the purposes of regulation. That such rate base continues to be called "value" should not confuse the well informed.¹

A modified concept of value is likewise used in taxation, especially in America, but unfortunately with less of candor than prevails among those who deal with public utility rate making. Examination of the property tax laws of the forty-eight states would show that these taxes are imposed universally upon the value of the taxable property and that, with few if any exceptions, the lawmakers had in mind the concept of value as used in economic science. The legislatures have gone out of their way to specify "market" value, "cash" value, "actual," "fair," "true and just" value, what the property would bring if sold "by a willing seller to a willing buyer," etc., terms which indicate their adherence to the economic concept of value. Property tax administration, as the result of circumstances which will be examined later, has not always succeeded in carrying out this idea of value, though still generally maintaining the pretense of a true valuation. Assessed values are very often not the same as true values, and they vary all the way from considerably above the true value to as little as five or ten per cent; as a rule assessments are below true value. It is safe to say that tax administration will never be competent to place true values upon all the complex kinds of property that are taxable in America. For certain classes of property it may become necessary to emulate the frankness of the rate-making authorities and set up a more or less arbitrary "tax-base" which will be workable.

The use of the term value in these varying meanings has been the source of much confusion — confusion which is to be escaped only by recognition that book values, rate-base values, tax values, etc., are generally neither the same thing as value in the economic sense, nor consistent with each other. It is not, as is sometimes assumed, indisputable evidence of fraud or corruption when a railroad reports one "value" to the rate-making commission and another to the tax assessor. Nor is there the supposed fine logic in the proposal

¹ See G. G. Tunell, *Value for Taxation and Value for Rate Making*, Proceedings of the Twentieth National Tax Conference, 1927, pp. 263-279.

to stop tax evasion by permitting the state to buy in any property at the valuation on the owner's tax return. Discrepancies between market and book values do not necessarily demonstrate either foolish investing or crooked accounting. And finally the situation does not demonstrate hopeless confusion as to the meaning of value.¹

EXERCISES

1. Suppose \$100 is deposited in a savings bank which pays three per cent interest on such deposits.

(a) What would this deposit be worth in three years' time if no money were withdrawn in the meantime?

(b) What would it be worth in six years?

(c) What would it be worth in fifteen years?

NOTE. Solve (a) exactly. For (b) and (c) work out the formula only.

2. Suppose the interest rate is six per cent. What is the present value of a promissory note for \$1,000 due in two years?

3. A stand of growing timber will be marketable three years hence at an estimated value over and above the cost of cutting of \$5,000. What is its present value if the interest rate is five per cent?

4. Suppose the interest rate is five per cent. What is the present value of a \$1,000 four per cent bond maturing three years hence? This description means that the owner of the bond will receive \$40 per year for three years and a redemption payment of \$1,000 at the end of the third year.

5. A certain piece of land can be made to return a perpetual net income of \$600 a year. What is its present value, (a) at a five per cent rate of interest? (b) at a four per cent rate? (c) at a six per cent rate?

6. In general, if income value remains unchanged what is the effect on the value of capital, (a) of a rise of the interest rate? (b) of a fall of the interest rate?

7. Draw diagrams illustrating the relation between the prices of competing goods.

8. Draw diagrams illustrating the relation between the prices of complementary goods.

9. Draw diagrams illustrating the relation between the prices of joint cost goods.

10. Draw diagrams illustrating the relation between the prices of "tandem" goods.

¹ For an interesting discussion of these matters, written from a somewhat different viewpoint, see L. D. Edie, *Economics: Principles and Problems*, 1932, especially pages 190-201.

SUGGESTIONS FOR FURTHER READING ON PART II, THE FORCES DETERMINING PRICE

Sections devoted to the subjects of price and value will be found in all of the general economics texts listed at the end of Part I. The following will be found especially useful:

- EDIE, L. D. *Economics: Principles and Problems*. Second edition. New York, 1932
- Recent Economic Changes in the United States*. Report of the Committee on Recent Economic Changes, of the President's Conference on Unemployment. New York, 1929
- FISHER, I. *Elementary Principles of Economics*. New York, 1912
- MARSHALL, A. *Principles of Economics*. Eighth edition. London, 1920
- MCCONNELL, D. W., and others. *Economic Behavior*. Revised edition. Boston, 1939
- MCISAAC, A. M., and SMITH, J. G. *Introduction to Economic Analysis*. Boston, 1937
- GARVER, F. B., and HANSEN, A. H. *Economics*. Boston, 1937
- MEYERS, A. L. *Elements of Modern Economics*. New York, 1937
- ROLL, E. *Elements of Economic Theory*. London, 1937
- SLICHTER, S. H. *Modern Economic Society*. New York, 1931
- TAYLOR, F. M. *Principles of Economics*. New York, 1921
- SHEARMAN, H. P. *Practical Economics*. New York, 1922

Among the works devoted more particularly to the subject of price and value, the following are suggested:

- NOURSE, E. G., and DRURY, H. B. *Industrial Price Policies and Economic Progress*. Washington, 1938
- MILLS, F. C. *The Behavior of Prices*. New York, 1927
- SCHULTZ, H. *Statistical Laws of Demand and Supply*. Chicago, 1928
- HENDERSON, H. D. *Supply and Demand*. New York, 1922
- WICKSELL, K. *Lectures on Political Economy*. London, 1935
- WICKSTEED, P. H. *Common Sense of Political Economy*. New York, 1910
- KNIGHT, F. H. *Risk, Uncertainty and Profit*. Boston, 1933
- FISHER, I. *The Theory of Interest*. New York, 1930
- BURNS, A. R. *The Decline of Competition*. New York, 1936
- PIGOU, A. C. *The Economics of Stationary States*. New York, 1935
- MEADE, J. E. *The Rate of Interest in a Progressive State*. New York, 1933
- PHELPS-BROWN, E. H. *The Framework of the Pricing System*. London, 1936
- CHAMBERLIN, E. *The Theory of Monopolistic Competition*. Cambridge, 1933
- ROBINSON, J. *The Economics of Imperfect Competition*. New York, 1933
- DAVENPORT, H. J. *The Economics of Enterprise*. New York, 1913
- MEEKER, J. E. *The Work of the Stock Exchange*. New York, 1922

See also the books on money and prices listed at the end of Part IV.

PART III
THE DISTRIBUTION OF WEALTH
AND INCOME

XVI

ECONOMIC RENT

The meaning of distribution. As used in economics the term "distribution" denotes the division of society's total income among the members of the social group. The study of distribution may be pursued with two different objects in view. We may set out to discover the actual incomes of the different individuals and to explain the inequalities which appear among them; this would be called a study of *personal distribution*. Or we may distinguish between *types* of income without reference to the total receipts of individuals, and investigate the forces which determine the size of these different types of income. This would lead us into a study of the different factors of production and of the economic laws which control their remuneration. Such an inquiry would be called a study of *functional distribution*.

We shall study both aspects of the problem of distribution in the following chapters, postponing our study of personal distribution until we have made a preliminary study of the functional types of income. There are four of these functional types: economic rent, interest, wages, and profits. Although there are similarities between these types of income when considered from certain points of view, each of them involves many problems peculiar to itself. Each type when viewed as a separate category of income is affected as to its amount by independent social forces. It is our purpose to examine these distinctive blocks of income separately, beginning with economic rent.

Nature and definition of economic rent. As used in the discussion which follows, *economic rent* is a technical term denoting the income which accrues to landowners from the use of their lands for productive purposes. At a given time a certain piece of land has power to contribute a definite, ascertainable quantity of income to the total social fund, over and above what could be produced by society's capital and labor without this land, and the owner can demand that

this quantity of income (or its value) be given to him as the condition upon which he will put his land to use. He has two ways of enforcing this demand. By holding his land idle, he can force others to bid for the privilege of using it in some business enterprise until the highest bidder offers him a money payment approximately equivalent to its income-producing power. Or he can work his plot of ground himself and thus derive directly the benefit of its productive power. In the first case, the owner of the land will receive a definite money payment, distinct from all his earnings as worker or capitalist, which should be called *contractual* or *explicit economic rent*. In the second case, he will receive an income of approximately the same value, but it will be merged in the total result of his activities as worker, capitalist, and landowner. In this case the portion of income traceable to the productive power of land should be called *implicit economic rent*.

It may be well at this point to recall that the term "land," as used in economics, embraces all the resources of nature afforded by man's physical environment, including the mineral deposits of the earth, the forests, the water power, the medicinal springs, the fisheries, etc. A further word of caution must be added with regard to those cases where economic rent takes the form of a contractual or explicit payment. Usually when land (or other natural resource) is leased by its owner, the lease covers articles of wealth imbedded in the land which are the result of man's labor and not the gift of nature. The tenant in the city for example commonly obtains the use of buildings, sidewalks, and other improvements in addition to the use of the site. In the country districts, fences, barns, drainage systems, orchards, go with the land. In both cases a single payment is made for the use of this composite unit of land and capital. This payment, from the standpoint of economic analysis, is made up of two different elements: *interest on capital* and *economic rent of land*. The study of economic rent is concerned only with that part of the landowner's income which may be attributed to the use of the land itself unimproved and devoid of capital.

From these principles, the following definition emerges. *Economic rent is a net income from the use of land, distinct from all income from labor or the use of capital.* To avoid ambiguity, the full term

economic rent will be employed when referring to this peculiar type of income. When required to refer to a contract payment which includes other things than economic rent, the word *rental* will be used.

Diminishing returns: In agriculture. A useful method of measuring the specific contribution of a given area of land is to imagine its withdrawal from use, thus compelling society to concentrate its productive effort on the remaining natural resources of the earth, and then to consider the effect of this change on the total income of society. If the withdrawal of a piece of land from use should cause no diminution of social income, it may be said that the land in question contributes nothing. Moreover on this assumption the owner of the land could not compel anyone to pay him an income for the privilege of using it. To the other members of the community it would be immaterial whether this particular acre were used or allowed to remain idle. But in general this supposition is contrary to fact; labor and capital cannot be withdrawn from a part of the land and concentrated on the rest without causing a diminution in the total income. The reason is that the exploitation of land and other resources is governed by the law of *diminishing returns*.

Let us take for example a plot of land, twenty-five acres in area, which is used for the production of oats. Assume at the outset that the entire area is tilled by the proprietor with his complement of seed, tools, and other capital, and that it produces 500 bushels of product. Because of the economy of time and effort made possible by division of labor, a second worker with an equal equipment of capital can, we may assume, double the product, making the total output 1,000 bushels. But this increase of the product by increasing the effort expended cannot go on indefinitely; the time will come when an additional unit of capital and labor will not increase the output proportionately. Let us suppose that this point is reached when a third worker and his equipment are employed. Whereas each of the first two workers with his equipment added 500 bushels to the output, the third man similarly equipped adds 400, thus making the total product 1,400 bushels. A fourth man with the same equipment adds 300 bushels, making the total 1,700. And a fifth man adds 200 bushels, making the total output of the farm,

when cultivated by five men, 1,900 bushels of oats. It will be seen that, although the total product of the land continues to increase with the increase of labor and capital applied to it, the *marginal* product — that is, the amount attributable to the last man and his capital equipment — decreases as the third, fourth, and fifth man are added. This declining marginal product is the symbol of diminishing returns. It is characteristic of all extractive industries. To be sure, the point at which diminishing returns will appear is not the same for all grades of land, for all sorts of crops, and for all kinds of agricultural organization; but sooner or later the point is reached where nature responds to increased intensiveness of cultivation with a diminishing marginal product.

The general principle of diminishing returns. In its most general form the law of diminishing returns may be stated as follows: *after a certain point has been reached in the exploitation of a productive factor of fixed amount, additional applications of other factors to its exploitation, though they will increase the total product, will not increase it in proportion to the increase of the variable factors.* As thus stated, it will be seen that the law applies when any one of the factors is fixed in amount while the others increase. A unit of capital of given size — say ten machines — may be operated at greater intensity by increasing the amount of labor and land combined with them; at some point in this process the output of the enterprise will begin to increase at a rate less than proportionate to the additions of labor and capital. Or if the labor supply is the fixed factor and capital and land the variables, the same phenomenon will appear: sooner or later successive additions to the increasing factors will cause less than proportionate increases in the total output. This law of diminishing returns is of utmost importance in determining economic rent in situations where land is the fixed factor. It applies to all kinds of land, but for purposes of illustration we shall continue with our simple example from agriculture.

What is implied in diminishing returns. So important is this principle, not only in the study of economic rent, but with regard to the other shares in the social income as well, that we may well pause to make sure of its implications. It should be noted, in the first place, that the statement of the existence of diminishing re-

turns does not imply that small-scale farming is more economical than large scale. It is a necessary postulate of this law that we deal with a given quantity of land, which quantity is held constant while the expenditure of labor and capital is increased. Comparison of the relative economy of small- and large-scale farming has to do with areas of different sizes. On none of these areas can the labor forces be expanded indefinitely without eventually reaching the point where the product increases less rapidly than the increase of labor and capital expended. The areas of various sizes differ from each other in respect of the operation of diminishing returns only with regard to the point at which the marginal product will begin to decline.

In the second place, the principle of diminishing returns must not be confused with the tendency of the soil to lose its fertility after successive years of cultivation. It holds true in the first year of cultivation of virgin land as much as in the results obtained after many years of tillage. In stating the principle of diminishing returns we are dealing with several simultaneous possibilities for a single season of cultivation, only one of which can actually occur. Referring again to our illustration, *in any given year* the twenty-five acre farm would produce 500 bushels if tilled by one man, 1,000 bushels if tilled by two, 1,400 if tilled by three, etc.

In the third place, in conformity with this implication of a given point in time, the principle of diminishing returns rests on the assumption that the technique of cultivation remains unchanged when successive units of capital and labor are applied to a constant area of land. The fact that improvements in agricultural machinery, fertilizing methods, and the science of tillage have made the output per worker larger now than formerly in no wise conflicts with the principle of diminishing returns. On each stage of the arts, from the beginnings of the rudest tillage to the modern stage of scientific agriculture, the marginal product of the land would have decreased if the number of workers per acre had expanded beyond a certain point.

Diminishing returns translated into increasing marginal cost. For purposes of further analysis let us throw our previous illustration into tabular form, as appears on the next page.

GIVEN: A FARM OF TWENTY-FIVE ACRES

<i>Number of workers equally supplied with capital</i>	<i>Output of the farm</i>	
	TOTAL PRODUCT IN BUSHELS	PRODUCT OF MARGINAL WORKERS IN BUSHELS
1	500	500
2	1,000	500
3	1,400	400
4	1,700	300
5	1,900	200

The third column of this table is the significant one. It records the marginal products of the workers; that is, the part of the total product which is attributable to the presence in the labor force of the last man who joins it. Diminishing returns set in with the employment of the third man. A three-man labor force with a proportional quantity of capital increases the total product from 1,000 to 1,400 bushels; hence the marginal product of three men is 400 bushels. Four men with their tools produce 300 bushels more than three men; this is the marginal product of a four-man labor force. Similarly it is shown that a five-man labor force produces a marginal product of 200 bushels.

It will be recalled that we have assumed each man in the working force to be equally well equipped with capital; on this assumption there is a uniform cost involved in employing each of the workers. If now we resolve the expanding working force into terms of monetary cost, it will appear that the principle of diminishing returns is but another expression of the *law of increasing marginal cost*, with which we have become familiar in our study of value. In doing this, care must be taken to include all the costs of cultivation, compensation for the labor and management of the proprietor, a normal interest return on the money invested in tools and equipment, and expenditures for labor, seed, and other materials. In the following table the matter is simplified somewhat artificially by assuming that all of these costs can be proportioned to the number of workers employed in cultivating the farm.¹

¹ The chief element of artificiality in our mathematical illustration obviously consists in the elimination of all costs in the nature of fixed charges. (See Chapter X.) Recognition of these costs, while complicating our analysis, would not in any way affect the conclusion. And as a matter of fact, in a simple enterprise such as this, fixed costs are of slight practical importance.

STATEMENT OF INCREASING MARGINAL COST

<i>Number of workers</i>	<i>Expenditure in dollars</i>	<i>Total product in bushels</i>	<i>Marginal product in bushels</i>	<i>Marginal cost per bushel</i>
1	\$100	500	500	\$.20
2	200	1,000	500	.20
3	300	1,400	400	.25
4	400	1,700	300	.33 $\frac{1}{3}$
5	500	1,900	200	.50

In the final column of the table is recorded the marginal cost of production. These figures are obtained by dividing the marginal products of the working force into that part of the total cost which was involved in the production of the marginal product. The first marginal product is 500 bushels; this product involves a cost of \$100; the marginal cost per bushel is therefore 20 cents. The marginal product of two men is also 500, produced by an additional expenditure of \$100; hence the marginal cost per bushel remains at 20 cents. The third marginal product, added by the expenditure of an additional \$100, is 400 bushels; the cost of this marginal product is accordingly 25 cents per bushel. By a similar computation the cost per bushel of the fourth and fifth marginal products is shown to be 33 $\frac{1}{3}$ and 50 cents respectively. We note again that the marginal cost of production rises after cultivation is pressed beyond the point of diminishing returns.

Marginal cost equated to market price. The bearing of this illustration on our main problem — economic rent — will become clear when we turn from a consideration of the costs of our farmer to take note of his receipts or income. Since he is producing goods for sale on the market, his money income will be determined by the price set upon his product through the operation of the forces of demand and supply. These forces determine that the price of any good produced under conditions of increasing cost tends to equal the marginal cost of production.

Our farmer for example will plan to produce 1,900 bushels of oats only in the expectation of a future selling price of 50 cents per bushel, or 1,700 bushels only if the price is expected to be 33 $\frac{1}{3}$ cents, or any other quantity only if the anticipated market price is high enough to cover the marginal cost of that quantity. This statement

in fact sums up the relationship of each individual producer to the market. Each assumes that the market price, at any given moment, is beyond his control and makes it his concern to avoid producing any part of his output at a loss. When he anticipates a rise in price, he is inclined to increase the application of labor and capital to the soil and thus advance his marginal cost toward equality with the new price, being induced to do so by the prospect of larger gains to be made from his increased output. Conversely when he expects a decline in price he tends to contract his production until he ceases to lose on the marginal units of his output. These are applications of the laws of value familiar to all students of the market. It should of course be recognized that the actual response of farmers to price changes is generally far more sluggish than our theoretical example might seem to indicate.

The concept of producer's surplus. All units of the same commodity sell in a given market at the same price. Since this price is equated to marginal cost, it follows that when production has passed the point of diminishing returns the sale of the total product results in a surplus return. The amount of this surplus, for any given quantity of product, may be discovered by comparing the money receipts from the sale of the product with the total cost of production. Assuming for example that the market price for oats is $33\frac{1}{3}$ cents per bushel and that our farmer produces 1,700 bushels, he will receive a gross money income of \$566.66. Reference to the table shows that his total cost is \$400; hence he receives a surplus gain of \$166.66.

The point should be emphasized that this part of the farmer's income is a *true surplus* over and above all payments required to induce production of the quantity offered for sale. It will be recalled that the total cost of production includes interest on all capital used, wages for all hired labor, replacement of all materials consumed, and a remuneration for the labor and management of the farmer adequate to call forth his effort. To avoid the appearance of having begged the question by juggling the conditions of our illustration, we should point out that no reasonable assumptions regarding this normal reward of the farmer will change the results of our analysis. We may take any figure we please as a measure of

this factor in the cost of production; it must still appear in our column of total costs. If the farmer's product sold for no more than enough to cover his total cost (including the normal reward for himself), it would nevertheless prove sufficient to justify his operation of the farm. The fact that it sells for more than this is to be attributed to the laws of nature, which cause man's exploitation of natural resources to proceed under the principle of diminishing returns, and to the common sense of producers, which impels them to grasp all the gain that is open to them and to avoid loss on any portion of their output when the remedy lies ready to their hand. This surplus is economic rent. A synonymous term for it, and one which fits exactly into the conditions of our illustration, is *producer's surplus*. It is inevitably involved in the income derived from the cultivation of the soil after production has been pressed beyond the point where diminishing returns set in.

Contractual economic rent equivalent to producer's surplus. Thus far we have assumed that the owner of the land is also one of the workers engaged in cultivating it, as well as the manager of the enterprise. The economic rent which accrues to him is an implicit income, not received as a separate and distinguishable payment, but mingled with his total receipts from the sale of his product.

Suppose now that our farmer wishes to retire from active labor and to that end offers his land for lease. He will of course expect that his earnings as a worker will stop. We may assume also that he turns his supply of agricultural capital — tools, machines, etc. — into cash and invests the proceeds in good bonds, thus continuing to receive that portion of his former income which consisted in interest. Obviously then his real financial sacrifice upon retiring from the land will be measured by the economic rent which formerly accrued to him. This accordingly is the price he will be disposed to ask from prospective tenants for the use of his land. It is also a price which the tenant will be willing to pay; for after subtracting this much from his gross income from the sale of his produce, the tenant will have left enough to pay all the costs of production and will retain for himself a normal reward for labor and management. Consequently the contract economic rent will tend under competitive conditions to equal the surplus gain implicit in the cultivation

of the soil by the owner. When the land has been leased on these terms, the owner will occupy an unusual position as a sharer in the social income. With no further expenditure of labor and rendering no service as a supplier of capital (or, as the case may be, with adequate compensation for these functions), he will continue to draw an income from his privileged position as a landowner.

It is true that our simplified analysis gives the transaction an appearance of precision and exactness which one does not find in the world of affairs. One can easily imagine cases of tenants who are willing to pay more than the economic rent in order to win a position of independence, or of landlords who consent to take less in preference to making extreme use of their bargaining power. Neither party to the lease contract can calculate the economic rent to a nicety, and when, as is normally the case, a certain amount of improvements goes with the land, it is beyond the power of human ingenuity to distinguish rent from interest to the last penny. Considerations of the probable future behavior of prices or of costs of production — considerations left out of account in our illustration because of our supposition of a fixed point in time — may also influence the calculations of tenant or landlord, though if these considerations affect the rental it will be only because they indicate a change in the economic rent for coming years. Such qualifications of our statement might be multiplied; but they do not alter the essential principle after all is said. Competition among landlords on the one hand and among tenants on the other works continuously to the end that contract economic rents approximate in value the producer's surplus inevitably resulting from exploitation of land under conditions of diminishing returns. Only when lands are leased for long periods at fixed rentals, so that the changing conditions of the market depart widely from the calculations of landlord and tenant, will the tenant be enabled to retain any considerable portion of the economic rent for himself or the landlord succeed in grasping more than this economic rent.

Cultivation of lands of different productivity. Our discussion thus far has dealt with a single piece of land of uniform quality per acre under cultivation by a single proprietor or tenant. If all the land of the earth were of one quality, rent would arise upon it as soon as the

amount of product demanded by consumers had grown so large that a portion of the supply must be produced under conditions of rising cost. But the arable lands of the world are not of uniform grade; differences in productivity exist because of differences in inherent fertility, in proximity to market, in facilities for acquiring the necessary supplies of capital and labor, and so on. These differences in quality, in this day of commercial agriculture, appear to the cultivator in the guise of varying costs of production on different plots of ground. The typical farmer in the United States usually has within the limits of his farm many plots of varying fertility, ranging from the choicest fields, where the product per unit of expenditure is high, to patches of barren soil that are allowed to lie waste. Similar gradations appear as between the holdings of different farmers. In each community there are certain lands whose nearness to market and abundant fertility afford the largest output with the least expenditure of effort, others less well situated or less fertile where the average cost of production is higher, still others far removed from the market or composed of thin or impoverished soils whose cultivation at existing prices of farm produce does not pay. As the next step in our analysis we need to take account of differences in the income-bearing powers of land.

Let us suppose, to make the case simple, that there are three different grades of soil, each of twenty-five acres and all devoted to the cultivation of oats. The farm of our previous illustration is Grade A, the most fertile of the three. The inferiority of Grade B soil as compared with Grade A is shown by the fact that the initial expenditure of capital and labor on B will return a smaller product than the same expenditure on A. Similarly the inferiority of Grade C to Grade B is shown by a still smaller product from the first expenditure of effort on this land. The relations of these different plots of ground to each other may be pictured in the tables on the following page, the first of which it will be noted is repeated from our previous illustration.

It will be seen that the marginal cost for the first units of output is lowest on Grade A soil, next lowest on Grade B, and highest on Grade C, and that the principle of diminishing returns applies to the cultivation of each grade of soil, though the point at which it begins

and the rapidity with which marginal costs rise as the product is increased are not the same for all.

GRADE A SOIL

<i>Expenditure</i>	<i>Total product</i>	<i>Marginal product</i>	<i>Marginal cost</i>
\$100	500	500	\$.20
200	1,000	500	.20
300	1,400	400	.25
400	1,700	300	.33 $\frac{1}{3}$
500	1,900	200	.50

GRADE B SOIL

<i>Expenditure</i>	<i>Total product</i>	<i>Marginal product</i>	<i>Marginal cost</i>
\$100	400	400	\$.25
200	700	300	.33 $\frac{1}{3}$
300	900	200	.50
400	1,000	100	1.00

GRADE C SOIL

<i>Expenditure</i>	<i>Total product</i>	<i>Marginal product</i>	<i>Marginal cost</i>
\$100	300	300	\$.33 $\frac{1}{3}$
200	500	200	.50
300	600	100	1.00

Economic rent on lands of different quality. Now it is apparent at first glance that, until cultivation has been pressed beyond the point at which diminishing returns appear on Grade A soil, the other two grades of land will not be cultivated at all, since it will not pay to cultivate them. Until the third expenditure of labor and capital has been made upon the land of highest quality and the total output of that land has been advanced to 1,400 bushels, the price of the product will not rise above 20 cents a bushel, this being the marginal cost. But the first units of produce on land of the second grade will cost 25 cents per bushel and on land of C grade, 33 $\frac{1}{3}$ cents; to produce them, when the market price is 20 cents, would therefore involve a loss to the farmer, which he will not bear. Up to this point (the third expenditure on Grade A) there will be no economic rent, even on lands of the highest quality, since total costs equal total income. If the price rises to 25 cents per bushel, it will then be profitable to make the third expenditure on Grade A soil,

raising its output to 1,400 bushels, and at the same time to make the first expenditure on Grade B, the total product from this second grade being 400 bushels. Grade C will remain out of use, for the market price is still not high enough to cover its initial marginal cost. At this point there will be economic rent on Grade A, where cultivation has passed the point of diminishing returns, but none on Grade B, where total costs are equal to total income. Suppose that the market price rises again, this time to $33\frac{1}{3}$ cents per bushel. A fourth expenditure will now be made on Grade A soil, a second on Grade B, while the first cultivation will be made on Grade C. Economic rent on Grade A will rise; on Grade B economic rent will appear for the first time; while Grade C land, though under cultivation, will return no producer's surplus.

The question may be raised: How can we assume that cultivation will automatically expand with each rise in price until the marginal units of the product merely replace their cost? Why does not the farmer on Grade A or Grade B soil restrict his output so as to be able to sell all the units at a price in excess of their costs? The answer is that he will either not have the power to do this, or if he has the power, it will be unprofitable for him to do it. If he is not the owner of the land, competition of other would-be tenants will cause a revision of the rent to keep pace with each increase in potential producers' surplus. The present tenant can retain possession of the land only by agreeing to bear these increased rent charges, after which, if he is to avoid loss, he will have no other alternative than to expand production to the point assumed. On the other hand, if he is the landowner it will be to his advantage to gain the maximum producers' surplus possible under prevailing market conditions, and this he can do only by advancing his marginal costs to equality with the price. Hence as the thickening of population causes an ever increasing demand for farm produce and a constantly rising price, the history of agriculture in the little society we have imagined will be marked by two lines of development. There will be a growing intensity of cultivation on the better grades of soil and a wider spread of cultivation to take in soils of poorer qualities. Along with these changes will go rising rents on all lands which were formerly rent bearing and a widening of the

rent-bearing area to include soils which formerly returned no producers' surplus to their cultivators.

Intensive and extensive margins. This illustration introduces a concept of importance in the doctrine of economic rent; the concept of the *margin of cultivation*, which may be defined as *that stage of exploitation at which a unit of expenditure just pays for itself*. This last, or marginal, expenditure is applied under such conditions that it produces an income only equal to the cost. That this is true may be due to either of two causes: the expenditure may be the last of a series applied to a given plot of ground, its product being restricted by the law of diminishing returns; or it may be the first expenditure made on lands so lacking in fertility or so remote from the market that, at the existing price of the product, they are barely worth cultivating at all. It is useful to distinguish these two cases by calling the one the *intensive*, and the other the *extensive*, margin of cultivation. The intensive margin of cultivation on any grade of land is reached when the product of the final unit of expenditure sells for no more than it costs to produce. The extensive margin appears on that grade of soil which it is just profitable to cultivate at a given time. Referring again to our illustration, with oats selling at $33\frac{1}{2}$ cents, Grade C soil lies on the extensive margin; the intensive margin is represented by the second expenditure on Grade B and the fourth on Grade A.

Since marginal cost equals market price, there is no economic rent derived from the expenditures on either the extensive or the intensive margin. As was made clear in our preceding illustration, the position of both margins at any given time depends upon the selling price of the product. If anything occurs to raise this price, the intensive margin on all lands under cultivation will be raised and the extensive margin will be pushed out to lands whose lack of productivity made them extra-marginal at the old price. A fall in the price will, on the other hand, cause a lowering of the intensive margin on all lands and a narrowing of the extensive margin by the process of throwing the poorest lands out of use. It should be noted that these statements hold good only on the assumption that costs of production remain unchanged as the selling price fluctuates.

Urban lands: Inequality. In the city, as in the country, we find lands devoted to different uses and unequally productive. Different sites vary in profitability from the point of view of any type of business that we might choose to consider. The retail stores, the theatres and hotels, the office buildings and apartment houses, the banks and wholesale establishments in any city of considerable size occupy sites which, even to the most casual observer, differ in respect of their adaptability to the purposes of the business. In most cities, this difference is so noticeable that it is not difficult to point out the enterpriser who has surpassed his competitors in the scramble for desirable locations. It is evident that each of these leaders in his own field would sacrifice income if he were forced to move to a less desirable location, even if he applied his labor and capital as abundantly and efficiently as before. There is no way to explain these differences in the income-producing powers of enterprises differently situated, when managerial ability and other factors are equal, except to attribute them to the relative productivity of the lands on which they stand.

Diminishing returns from urban lands. Diminishing returns on urban lands, for many lines of use to which they may be devoted, may not be apparent until utilization of the land has reached a point of high intensity; but the principle is ever present. For example in retail trade, an increased expenditure of capital and labor on a site of given dimensions takes the form of additional stories to the building, a larger area of floor space, an increased stock in trade, increased working force, a more complex and unwieldy organization. As the building grows in size, the costs of maintenance, of transportation of goods and passengers from floor to floor, of supervision, will eventually rise. Many small sources of waste will pass unnoticed by the managers which in a smaller establishment would be discovered and corrected. For many reasons the costs of rendering the service will rise, or to use the correlative expression, the returns from the last unit of expenditure will fall. Hence the time must come when additional expenditure on the best sites no longer affords a larger profit than the same expenditure on lands not so well situated.

The exploitation of each parcel of land tends to be pushed to the

limit — typified by an added story, department, operating device — which it is worth while to operate at a given time on any particular site of given size. This is the intensive margin. The extensive margin of land use is not to be found in retail trading or in any other use of city land. There is usually no no-rent land within the limits of the ordinary city. What occurs is that the business in question, as it expands, pushes constantly out from the focus of the most desirable land, taking in locations successively less desirable till it reaches those parcels of land which are no more useful for this business than for other uses. Beyond this the business cannot at any given time go, because there now appear other uses for the land which will yield a higher rent, this rent being determined by the net income to be derived from some other use of the land and resting finally upon the prices obtainable for the services of the land in this other — most productive — use. There is equality of return between the intensive margins on all the different parcels of land devoted, at any given time, to the particular business, and as conditions change this equality tends to maintain itself. With increased demand for retail services, for example, the intensive margin is raised by more complete — and more costly — utilization of sites already in use, while at the same time the business takes over certain sites which were formerly more profitably devoted to other uses but which the increased returns of the retail business now make it possible to wrest away through the offer of higher rents.

Various qualities, uses, and rents of land. The earth's surface thus contains lands of all degrees of productivity, from the so far useless desert tracts to the fabulously valuable office sites on Wall Street, New York City, or State Street, Chicago. Should we try to visualize this, we would have a panorama something like this: Beginning with the poorest land, we see the vast areas of millions of acres of sub-marginal lands — lands which thus far have proved incapable of any service to mankind. Next we come to the lands of lowest productive capacity, devoted to ranges for sheep and cattle, or — a little higher in the scale — to enclosed pasturage. Certain lands devoted to forest-growing are close to the margin. Then come lands employed in the raising of wheat, corn, and other

grains, cotton, and tobacco. Higher in the scale come the lands nearer the cities, raising celery and lettuce and other market garden crops. Soon we come to small town lots upon which are located modest retail stores, simple industries such as shoe repairing, blacksmithing, and laundering; followed by drug stores, restaurants, factories, garages, etc. We are now getting into the cities, having passed by the way lands used as golf courses, race tracks, bathing beaches, and so on. All along the line are residence lots, from the simple farmstead, on land little if any differentiated from the cultivated fields, up to the exclusive country estates and the costly sites of elaborate city homes. As we go from the small city to the large, and from the periphery toward the centre of the large urban districts, we pass in order a great variety of lands devoted to separate homes and apartment houses, retail stores, manufacturing, theatres, hotels, office buildings, and banks.

The foregoing is of course a most sketchy account of the panorama as it actually is. It will be sufficient however to remind us of certain fundamental facts. In the first place, we note that land is not uniform in quality nor all devoted to the same use or uses. Variety in quality arises obviously from the physical character of land. But it comes also from the position of land with respect to human habitation. Intensive settlement at any place, such as a large city, gives a quality to the land which is not possessed by lands of the same physical character in a sparsely settled region. We observe further that the different uses to which the various parcels of land are devoted result in products per acre of different value in a wide scale, from the marginal farm land, yielding no surplus above cost, up to the business site in the heart of a great city, bestowing a yearly fortune upon its owner.

It may help us to visualize economic rent as a surplus if we conceive of all the lands in the United States arranged in a scale of productivity from the worthless sub-marginal desert up to the most desired business site in New York City. The poorest land devoted to raising oats which we considered in our first illustrative example was no-rent or marginal land; had it been any less productive it would have been of no use. Other lands in our illustration yielded surpluses (economic rent) of various amounts according to their

utility for raising oats. But by going a bit closer to the city, we come to lands capable of yielding a surplus greater than that of the best oat land by being used to grow celery. Now oats could, without doubt, have been grown on the celery lands. But the utmost product of oats would yield a surplus less than is obtainable from raising celery on lands suited thereto. Further on is a residence section, where the economic rent per acre is greater than that from the best of the oat land or the celery land. And so on, through successive grades of land and constantly increasing economic rent, to the end of the scale. On each parcel of land superior to the no-rent grade there is produced economic rent, a surplus indicated in each case by comparison with no-rent or sub-marginal land at the bottom of the scale.

The value of land. We have established the principle that each landowner tends to receive an implicit or contractual income equal to the contribution which his particular plot of ground makes to the social income. The extent to which society suffers loss when a piece of land, urban or rural, is withdrawn from use is not measured by the total former product of the land, since holding this particular plot idle will set free labor and capital productive of income in other uses. This labor and capital however must now be applied at the margin, either on lands too poor to be utilized at the time or as the least effective units of expenditure on better grade lands. Consequently this capital and labor will be less productive than it was before freed by the action of the landowner in question. *The difference between the total product of a given piece of land and the product resulting from an equal expenditure on the margin of cultivation is the measure of the specific contribution of that land to the social income.* This is what society would lose if the land were withdrawn from cultivation by its owner. But it is also the measure of the economic rent of the land in question, as we have seen in the preceding discussion. Here then we have two equivalent expressions for the measure of economic rent, one viewing the matter from the social point of view and stating the measure of rent in terms of product, the other, from the point of view of the private individual, measuring economic rent in money terms.

It is this specific income from land — the economic rent — which

together with the interest rate determines land values. At an earlier point in our study¹ we considered the discount process as it applies to the valuation of durable capital instruments. It was there observed that, whereas the demand schedules for capital goods are always derived by discounting expected future income, cost of production operating on the supply side is also a factor in the value of such categories of these goods as are reproducible. In the case of capital instruments which cannot be reproduced, it is the discounting of future income which alone operates on both demand and supply to determine value. With trifling exceptions of insufficient moment to affect the general rule, land is to be classified with this group of goods. Its value varies directly with economic rent and inversely with the interest rate. This is not to say that the sale price of any piece of land is determined by the present rent capitalized at the present rate of interest. Considerations as to probable future changes in rent and the rate of interest always enter into the calculations of the present value of future income; but alterations of land values resulting from these considerations are all in harmony with the basic principle, which may be stated as follows: *the value of land (and other natural agents) is derived, through demand and supply, from buyers' and sellers' estimates of the discounted value of expected economic rent.*

Relation of rent to prices. There is a good deal of popular confusion regarding the causal relation existing between the rent of land and the sale price of the product of the land, it being commonly believed that high rents cause high prices. But this popular notion confuses cause and effect; it is the price of the product which affects the economic rent of land, not the reverse. In the case of rural lands the true relationship between rent and the price of the product becomes quite clear when we review our previous analysis. The price-determining costs of production in agriculture are the marginal costs. But on the margins, as shown in our preceding discussion, cost is equated with price, allowing no producer's surplus or economic rent from the sale of these marginal units of supply. In other words, economic rent is not included among the price-determining marginal costs. Surplus receipts are obtained from the

¹ Pp. 331 f.

sale of the units of supply other than marginal because the existing market price makes production profitable beyond the point of diminishing returns. If rents are high at any given time it is because prices are high.

As to urban lands, the fallacious belief that high rents cause high prices may be traced to a misconception of the nature of the product of these lands. The product of land devoted to retail trading is obviously not the articles offered for sale in the shops. It is an intangible service, the service rendered by the merchant to the social group by promoting the ease and convenience of supplying human wants. If we wish to speak of the price of this service, it should be clear that we refer to the reward of the merchant, not the sale price of his stock in trade. Any attributes of a site which increase the merchant's gain on a given expenditure of capital and labor in the retail trade beyond the amount received by the users of marginal land will bestow a rent on the site. In some cases a given site will increase the merchant's profit for no other reason than that it places him in a position to overcharge his patrons. Here there is a relation between rent and the selling price of goods, but the causal connection runs from price to rent, not from rent to price. The immediate cause of the high rent, as always, is the large return of the enterpriser; the high return is caused in this case by high prices.

But the selling price of his goods is only one of the factors affecting the enterpriser's net income; another factor of equal importance is his cost of operation. If a merchant has low operating cost because of the character of the site upon which his shop stands, it is the land which really produces his extra profit, and the owner of the land will absorb the excess in the form of a high rent. Among mercantile enterprises large volume of turnover in comparison with necessary overhead expense is a primary cause of relatively low cost of operation. Sites on crowded thoroughfares and at the transfer points of traction lines permit a volume of business so much in excess of that done by shops in less effective locations that the merchants who possess them can gain large returns at prices no higher than those charged by their competitors. In such cases high rents will be associated with low selling prices of goods; though it is

still correct to say that the high rent is conjoined with a high reward for the merchant's service; *i.e.*, with a high income or gain.

The point to be emphasized then is that high rent is caused by those attributes of the site which enable its possessor to derive a large income from its use. Ability to charge high prices is only one of these attributes, and a relatively unimportant one. The buyer in any city will find that low prices are as frequently to be found on high rent lands as are high prices. In fact it is upon these lands that the so-called "cut rate" shops flourish. The United Cigar Stores, for example, are located only at the corners of busy streets; the five-and-ten-cent stores, the chain drug stores, the large department stores, and other enterprises which advertise a low price policy are similarly located on expensive land. Our illustrations have been drawn from one type of enterprise, retail trading, but the student will have no difficulty in applying the principle they embody to lands in other uses.

Quasi-rent. It will be recalled that the principle of diminishing returns, as we have stated it, is applicable under appropriate conditions to the exploitation of any factor of production. The condition essential to diminishing returns is that one of the factors be held constant while those which coöperate with it are increased. As the variable factors increase in amount the output of the combination, while increasing absolutely, will diminish as measured in proportion to the increasing factors. As we shall see in subsequent chapters this principle is of great importance in determining the other functional types of income — interest, wages, and profits — as well as rent.

We should note at this point that whenever a situation arises in which either labor or capital is a fixed factor, with the other factors variable, its income will be of the same nature as the economic rent of land. A labor group, if absolutely fixed in number and indispensable to productive enterprise, would derive a producer's surplus determined by forces analogous to those discussed in the preceding pages. Similarly the owner of capital instruments, if their supply cannot be increased and they are in great demand for productive purposes, can exact a payment of this kind for their use. Ordinarily the competitive forces of the market fix the rates of

return for labor and capital in a quite different way, because ordinarily and in the long run the supplies of these factors can be and are increased in amount. It should be carried in mind however that when a segment of the labor or capital supply is placed in favoring circumstances — limited in amount and combined with increasing supplies of the other factors — its income will diverge from the returns normally allocated to that factor through the market and become a *quasi-rent*.

Economic rent distinguished from other income. The condition of fixed supply is much more important in determining the income of land than that of the other factors. There is no productive factor, other than land, which cannot be increased in amount over long-run periods, either in its entirety or in its special forms. The whole mass of the laboring population can be, and normally is, increased by an expansion of the birth rate, a fall of the death rate, or both; and as for the fractions of the whole labor supply which make up the different occupational groups, they expand and contract under the influence of training and education. Continuous redistribution of the labor supply tends to prevent one group from long enjoying a differential advantage as compared with other groups.

This is even more true of capital instruments. Through the voluntary savings of its members, society's total equipment of these instruments is increasing as the years pass and may be increased at any given moment by making special appeals to the savers or offering special rewards to investors. No particular type of capital instruments can be held permanently fixed in supply. When the income enjoyed by the owner of a particular machine begins to outstrip that of owners of other types of capital instruments, equal in cost of production, new machines of this type will be made and set up in competition with those now in existence. Capital flows from one form to another, seeking out the points of maximum income. This flow may be delayed or retarded by special monopolistic devices, such as patents, secret processes, and copyrights, but it tends constantly to work out an equalization of reward to owners of capital instruments proportionate to the cost of these instruments.

This principle and the fundamental difference between land and

capital instruments which it expresses may be illustrated by a simple example. Suppose that an office building is erected at a cost of \$100,000 upon a certain city lot. A year's operation shows the following results: total income, \$15,000; operating costs, \$5,000; interest on capital (at 5%), \$5,000; net income, \$5,000. This last item is true producers' surplus: it is, in fact, economic rent, being due solely to location on this particular piece of land. So long as conditions generally remain unchanged, it will persist.

Now let us consider a capital instrument in precisely the same position. A shrewd enterpriser builds, at a cost of \$100,000, a ship to carry passengers from the city to a nearby shore resort. The results of his first year's operations are exactly the same as those of the office building; *i.e.*, total income, \$15,000; operating costs, \$5,000; interest on capital, \$5,000; net income, \$5,000. Here also is a surplus income, the reward of superior business foresight. But it does not rest on the firm foundation of economic rent. Very soon other enterprisers, seeing the extra profits offered, will build other ships. Competition will reduce the rates charged passengers; the gross income per ship will decline in consequence; and finally there will accrue to the owner only the \$10,000 necessary to cover operating costs and interest on his capital investment. Thus does competition tend to make short-lived the surplus gain arising on occasion from shrewd applications of capital.

As with capital, so also among enterprisers. Access to this position of peculiar independence in the labor world is thrown open through society's educational system to all who possess the requisite ability. And the enterprisers vie with each other for the possession of managerial methods and of lines of business undertaking which promise unusually large rewards of income.

"Unearned increment." Economic rent as a private income is peculiarly hard to justify on grounds consonant with the common man's impulsive sense of justice and fair play. Economic rent is not a payment for any service rendered by the recipient, for it exists only when the recipient's income is in excess of adequate remuneration for all his labor as a worker, all his service as a saver and investor of capital, all his utility to society as an enterpriser. A service, it is true, is rendered in exchange for economic rent, but it is

rendered by nature, not by the rent receiver. In other words society, in permitting the existence of economic rent as a private income, is making payment for a service which could be had for nothing.

Because rent is of this character it has often been described — and condemned — as an unearned increment. As the words imply, this name refers particularly to the tendency of rent to increase as population grows. But there is no reason in principle why distinctions should be drawn between different cases of economic rent. Whether rising, falling, or remaining constant, all true economic rent is, in the common meaning of the term, “unearned” by the landowner; it is an addition or increment to his income in excess of adequate payment for every productive service he performs.

It is of course true that the present recipient of economic rent may have invested his savings in the purchase of the land and is therefore receiving only a justifiable return on his investment. For example assume that someone pays \$40,000 for a site which returns an economic rent of \$2,000. Clearly there will be no unearned increment in the income of this individual — at least until the economic rent increases — for the yearly rent represents only the normal return on his financial outlay. But the former owner is still receiving \$2,000 yearly from the investment of his receipts from the sale of the land, and if we agree that part or all of this income was unearned increment before the sale, it is still unearned so far as he is concerned.

Attacks on economic rent. Rebelling against the apparent injustice of the economic rent income, people with a bent toward social reform have frequently proposed that this income be taken out of private hands. These programs of reform have taken many different forms, from outright nationalization of land, as an extreme example, to milder proposals to appropriate a part or all of the economic rent for social purposes. The method suggested for such appropriation of rent is usually taxation; and this proposal too has sometimes taken the extreme form — as in the case of the single tax movement — of demanding that the state tax all landowners the full amount of the economic rent. Such a tax policy would turn into a common treasury the entire surplus income from

the use of natural agents. This income could then be shared among all members of society, either through expenditure by the state upon projects of general social utility, such as education, public health, alleviation of distress, or in the form of relief from the burden of other kinds of taxes. The strongest argument for the tax on economic rent is that it would produce a large revenue which might be used for the advancement of general social well-being, without deterring economic enterprise. This argument is based on the nature of rent itself. If economic rent is truly a surplus income, it can be withdrawn from its recipient without encroaching on those elements of his income — wages, interest, profits — which induce him to put forth productive effort. Our understanding of the theory of economic rent will make it clear that when the matter is considered in the abstract and at a given instant of time this argument is valid.

The arguments against the single tax are of a practical character. It is always difficult and frequently impossible to distinguish the economic rent income with precision from the other types of income in a business enterprise. The tax might easily encroach on these other types of income and thus act as a deterrent of productive effort. Moreover lands have changed hands so often that many present owners have paid full value for the right to receive economic rent. The unearned increment in these cases is accruing to people who at the present time are not landowners at all. It is obviously impossible to discover for the purposes of taxation who these people are, but also obviously unjust to allow them to escape the tax and load it on the man who has bought from them the rent income. To press forward a reform of this kind, destroying private property in land, might, owing to failure to grasp the distinction between land and other capital goods, launch a movement for wholesale abolition of all private property. However impressive the abstract arguments for the single tax, the practical difficulties in its way and the portentous character of its possible indirect effects make this reform one that is not likely to receive wide approval.

EXERCISES

1. The figures below show the results of successive added expenditures on different grades of land devoted to the production of wheat:

<i>Grade A Land</i>		<i>Grade B Land</i>		<i>Grade C Land</i>	
ADDED EXPENDITURE	ADDED PRODUCT	ADDED EXPENDITURE	ADDED PRODUCT	ADDED EXPENDITURE	ADDED PRODUCT
\$100	200 bu.	\$100	100 bu.	\$100	80 bu.
100	150	100	90	100	70
100	125	100	80	100	60
100	100	100	70	100	50
100	80	100	60		
100	60	100	50		

- (a) Convert the figures for diminishing physical product with additional outlays on different grades of land into figures expressing rising marginal cost.
 - (b) Plot the curves of marginal cost for the production of wheat on the three grades of land (on one chart).
 - (c) How far would cultivation be carried on each grade of land, if the market price of wheat were \$1.00 per bushel? if the market price were \$1.25 per bushel?
 - (d) Locate the intensive and extensive margins of cultivation when the price of wheat is \$1.00 per bushel; when the price is \$1.25.
2. A piece of land occupied by a warehouse is leased for \$12,000 per year. Taxes, insurance, maintenance, and depreciation consume \$5,000 per year. The building cost \$100,000. The rate of interest is 5%. Explain the two elements of which the owner's net income is composed, and determine the amount of each.
3. With the construction of the Erie Canal, the price of wheat at New York City declined. How did this affect the intensive and extensive margins of cultivation in New York and New England and the rent of land devoted to wheat production in this area?
4. Explain how each of the following factors may influence the use of urban sites:
- (a) Convergence of subway lines
 - (b) Proximity to schools
 - (c) Easy access to railroad terminals
 - (d) Freedom from noise
 - (e) Proximity to tidewater
 - (f) Migration of population to suburban areas
5. (a) How would the movement of the New York Stock Exchange to New Jersey affect rents in the New York financial district?
- (b) How would the replacement of an elevated street car line by an underground transit line affect rents and the use of urban sites along the route?
- (c) How would the development of the St. Lawrence Waterway affect urban rents in Boston? in Chicago? in Baltimore?

XVII

INTEREST

The owners of capital, like the owners of land, receive a share of the social income in excess of their current earnings as workers. This type of income is known as *interest*. Although similar to rent in that it accrues to its recipient by virtue of his ownership of wealth, interest is governed by different laws. It is our purpose in the present chapter to explain why interest exists as a separate type of income and to examine the forces which determine its amount.

Time preference. It is an attribute of human nature to be impatient for present enjoyment. It is in man's nature to postpone enjoyment only with reluctance and with a sense of sacrifice. Today's goods are in general preferred to future goods of like kind and amount. Conversely one naturally feels a distinct sense of gain when, having reconciled himself to waiting for a pleasure which has been promised in the future, he is enabled to abbreviate the period of waiting. Although the pleasure is the same in magnitude, with its nearer approach it grows in desirability and is more highly prized. This being true, identical goods when separated by intervals of time do not have identical present desirabilities. From this human trait there appears what is called *time preference*, or *the inclination to place a higher estimate upon present goods than upon identical goods in the future*.

While it is generally true that people place a higher estimate upon present than upon future goods, the degree of this time preference is by no means uniform. Each individual's time preference depends upon a number of factors, residing in his personal character and in the particular external circumstances in which he is at the moment placed. It occasionally happens that a certain person will actually prefer future goods over identical present goods, in which case we say that his time preference is negative. One to whom this time element is a matter of indifference is said to have zero time

preference. We may well note a few of the considerations that affect individual time preference.

Some men are by nature so impatient and impulsive, or by training so undeveloped in foresight and self-control, that present desires are unrestrained by the thought of future needs. Other individuals, though more stable in disposition, lack ambition and have on this account little motive for thrift and abstinence. All such people may be described as having a very pronounced positive time preference.

Love and a sense of duty toward dependents, on the other hand, tend to lessen positive time preference and may even bring about, at least beyond a certain point, a negative time preference. Thus a man may stint himself to buy life insurance for the sake of his family.¹ It is obvious however that not every man responds to this call of affection and duty.

A person's age and situation in life have their effect on his time preference. A young man before marriage generally has a pronounced positive time preference. After marriage this tendency may be very much modified. Never having saved before, the young man may begin to provide for the future with great concern. But he may also be so absorbed in meeting increasing demands on his income that provision for the future seems impossible. An older man, feeling life slip by, may for that very reason develop a very high degree of positive time preference. On the other hand, he may be so interested in the future welfare of his dependents as to manifest beyond a certain limit a lower or even a negative time preference.

The size and timing of people's incomes have a very important influence on their time preference. Individuals with large and stable incomes tend to have lower time preference than those with small or irregular incomes. Poor people generally have very pronounced positive time preference. When present income merely provides a comfortless existence, future needs are often forgotten. The size of anticipated income is very important. The man who

¹ A life insurance policy is not necessarily an indication of negative time preference, for even in life insurance interest is paid. It may be at a low rate, but it indicates nevertheless the general prevalence of some degree of positive time preference even among people who take out insurance.

expects an inheritance or a future return on an important investment has less reason to curtail present expenditure. On the other hand, he who foresees the end of his own earnings as a productive worker or the lessening of his income as an owner of wealth is likely to be impelled to save.

It is sometimes convenient to be able to express time preference as a rate. This matter is precisely analogous to the concept of rate of interest which was developed in Chapter XV. Thus if a person at any given time has such time preference that \$100 of income today would be to him of the same value as \$106 a year hence, we say that his rate of time preference is $\frac{6}{100}$ or 6 per cent.

Time preference and interest: Production of capital requires time. A period of time intervenes between the cost involved in the construction of every capital instrument and the recovery of this cost through the sale of its product. In many cases this period of time is of long duration. A new railroad for example is a group of capital instruments whose initial cost lies far back in the past, when the first steps were taken to obtain the raw materials which eventually took the form of its roadbed and rolling stock. Income from its use is realized only after the lapse of many years. With other capital instruments this time element may be insignificant. The production of a simple tool may require only a very short time; it may begin to yield consumable income almost immediately after the first costs involved in its production are borne. But whether of longer or shorter duration, the time element involved in the supplying of capital instruments is never entirely absent. This is a matter with which the student has already grown familiar through his study of the roundabout or capitalistic mode of production.¹

Interest a payment for waiting. Capital formation then requires waiting for consumable income, and time preference makes such waiting disagreeable to mankind in general. This means that the suppliers of capital instruments must bear two different kinds of cost: the ordinary cost of production and the cost of waiting for deferred returns. There must be remuneration for both of these costs if capital instruments are to be produced. That this

¹ See especially Chapters II and XV.

is true of the ordinary costs of production requires no explanation. To make it clear that there must also be remuneration for the cost of waiting let us consider on what terms the ordinary man will exchange present goods for future goods. Obviously he will not voluntarily engage in the transaction unless the future return appears at the present moment to be of desirability equal to the present outlay. But since the dollar in hand and the dollar to be received one year hence do not have equal desirabilities when viewed from the present, the two will not be exchanged for each other unless something in the nature of a surplus or premium is first added to the future dollar. This surplus or premium is *interest*, the reward for waiting. As long as waiting is irksome and capital accumulation is done voluntarily, society cannot obtain an adequate supply of capital instruments unless the suppliers are allowed a share of the social income which includes this net return in excess of the ordinary cost of producing capital.

The universality of interest. We see that interest is involved in every exchange of present for future satisfactions. But there is no form of activity open to man whose return does not lie in the future. The building of a house today is rewarded by its shelter in the years to come; the effort expended in planting a tree is rewarded by the fruits borne in the future; the preparation of today's dinner is labor endured sometime in advance of its reward; at least a moment's time separates the filling of a pipe with tobacco from the satisfaction enjoyed by the smoker. But although we cannot change this order of nature, we do have some choice as to the degree of futurity of our income. Effort expended today may be rewarded in the immediate or in the remote future according to the direction taken by the expenditure. A day's labor may be consumed in tilling the ground in preparation for a harvest some months hence, or in digging an irrigation ditch whose benefits lie further in the future, or in repairing an automobile in anticipation of a pleasure ride tomorrow. There is an element of interest implied in all these activities; when the individual has to choose between two or more lines of expenditure whose incomes differ in futurity, his degree of time preference will act as a force determining the date at which satisfactions shall be seized.

Nature of the loan contract. In the case of many economic activities the interest element does not appear as a separate and distinguishable payment. When a capital instrument is used by the owner, the interest element, though present, may be lost from sight in the composite income stream. It is in connection with business transactions in which capital instruments or their value in money are loaned that interest appears in a form subject to measurement.

When we analyze the ordinary loan contract we discover that it represents a transaction in which present and future incomes are exchanged against each other. No borrower desires money for its own sake. When the business man borrows it is with the intention of adding to his factory equipment, enlarging his stock of raw materials, setting labor to work upon a productive process; in short, the real reason for borrowing is to acquire goods or services in excess of his present means of payment. The man who borrows for immediate consumption has a similar purpose in mind. The money he borrows will be used to finance a pleasure trip, to pay a doctor's bill, to buy food, clothing, or furniture. With him too the real purpose is to acquire certain kinds of wealth or service in excess of his present means of payment. Behind the borrowing of governments lies a similar motive. The Government of the United States obtains from the buyers of its bonds the purchasing power needed to buy munitions, clothing, or food for the prosecution of a war, or the wealth and labor required to build a canal or irrigation dam; the states borrow to build roads; the cities to erect school buildings, pave the streets, develop sewer or water systems. In all these cases the real wants of the borrower are satisfied only when the money advanced by the lender has been exchanged for wealth or services in certain forms. The effect of the loan from the borrower's standpoint is to increase his present stock of income-yielding goods.

Upon the lender the transaction exerts an immediate effect which is the opposite of this. He transfers to the borrower a part of his present means to acquire real income. Whether he deposits his purchasing power in a savings bank which uses it as a basis for a loan to a client, or gives it to a friend in exchange for his I O U, or

transfers it to a corporation as the purchase price of a bond or share of stock, or buys a government certificate of indebtedness is, after all, immaterial from the standpoint of the immediate effect of the transaction upon himself. In any case this typical lender must make an immediate sacrifice of an amount measured by the desirability of the goods which he might have enjoyed if he had chosen to spend his money rather than invest it.

It is apparent that, when the loan is repaid at maturity, a reverse transfer of goods takes place between borrower and lender. Viewing the transaction in its entirety, we see that the lender has given up present income in exchange for future income, and that the borrower has made the opposite kind of exchange, acquiring present income at a sacrifice of future income. This is the essential feature of all loan contracts.

The loan market and the rate of interest. When we view loan transactions in this light it becomes apparent that interest is a payment for the service of waiting and that the rate of interest is the price paid for that service. The rate of interest is normally expressed as a percentage; the ratio of the amount of one year's interest in terms of dollars to the amount of capital also in terms of dollars. Thus a borrower receiving \$1,000 may contract to repay the same amount one year hence and in addition to pay \$50 interest. The price of the loan — or the rate of interest — is $\frac{50}{1000}$ or 5 per cent per annum.

Like any other price, the rate of interest is determined in the market by the forces of demand and supply. The modern world has evolved an organized market, in which exchanges of present goods for future goods are constantly taking place; *viz.*, the loan or investment market. Because of the vast number of transactions which take place in the loan market and the innumerable ramifications of this market through the business world, the rate of interest there determined becomes the governor of the terms on which all types of transactions involving an exchange of present for future income occur. Further inquiry into the problem of the rate of interest necessitates an analysis of the circumstances affecting the demand and supply of loans. We begin with the supply side of the market.

The supply of loans: Reactions of individual lenders. Lenders react in general to the rate of interest as do the suppliers of any other service or commodity to its price. Amounts available for loans conform with the general law of supply. Under any given set of circumstances they vary directly with the rate of interest. The supply of loans however is subject to certain special peculiarities. Many people would postpone the enjoyment of some part of their wealth even though they received no interest. Men will make provision for old age, for the care of their children and other dependents, for unpredictable contingencies, such as accident, sickness, or premature death. Such savers accept interest on personal loans, savings bank deposits, mortgages, bonds, etc.; but even if these opportunities were not open to them, they would still save out of present income. Many well-to-do people habitually carry checking accounts at the banks in excess of their needs for the sake of the security and ease which a large balance affords them. Many small depositors likewise continue to maintain their accounts even though service charges are levied against them. The balances carried in commercial banks by both these classes of depositors become the basis of loans to business enterprise. It is evident therefore that some postponement of present consumption and some consequent formation of capital would occur, not only if no interest were paid, but even if a charge were made for the privilege of saving. In fact to some extent such waiting actually does take place.

Saving, it should be noted, has never been in its entirety a matter of cold calculation. With most people this is a rough and irregular process. Sometimes a person, knowing the variability of his own disposition, takes out an insurance policy in order to force himself to save. Individual saving on the whole is governed more perhaps by custom, habit, and emotion than by rational thinking. Saving or not saving is likely to be more a matter of habituary action than of logical decision. The proverbial thrift of the Scotsman is an inherited tradition. Any average individual's notion of what it is possible or proper to save is not so much the result of his own original thinking as of the attitude prevailing in the group of people among whom he was brought up or with whom he now associates. However in the psychology of groups as well as of individuals

there is still a rough balancing of marginal utilities, and certainly there is no reason to believe that the rate of interest is entirely without influence on the volume of saving and the amounts available for loans.

Corporate saving. It is also important to note that certain developments in our economic organization are tending to compel men to save, either against their will or without their knowledge, and thus to create a supply of loan funds without reference to the interest rate. One important tendency of this order is an outgrowth of the corporate form of organization, especially in the large-scale enterprises so common in our modern business world. Immediate control of the policy of a corporation lies in the hands of the directors, who are in turn responsible for their actions to the stockholders who elect them. When the corporation has grown to immense size and its stock has become widely scattered among a large number of small investors whose knowledge of the business is vague, the directors are virtually given a free hand in shaping the policy of the undertaking. One of the problems which they decide relates to the disposition of the corporation's net income. This belongs to the stockholders; it constitutes interest and profits upon advances made by them. But it is never the practice of prudent directors to distribute the net income in the form of dividends as rapidly as it accrues. A part is usually set aside as a reserve against future contingencies which may imperil the corporation's ability to pay its debts; another part may be devoted to improving and expanding the material equipment of the business; still another, to experimentation with new processes and selling methods. In thus disposing of the corporation's income, the directors make deductions from the immediate incomes of the stockholders, more or less without consulting their wishes in the matter, and add to the capital fund of society. An examination of the capital accounts of large corporations over a period of years will bring to light immense annual additions to capital equipment resulting from this practice of deducting investment funds from the incomes of the people in advance of the receipt of the income by its owners. Between 1910 and 1920, according to the estimates of one statistician, incorporated enterprises in the United States reinvested on the

average two billion dollars a year. A more recent estimate made by Berle and Means¹ indicates that, in the six-year period from 1922 to 1927 inclusive, 108 corporations (out of the 200 largest for which consolidated statements could be obtained for each year) saved 38.5 per cent of their net income available for dividends and that in the same period all corporations combined saved 29.4 per cent of their net income. It is of course impossible to say what proportion of these savings would have been made in any case, if the income of the corporations had been distributed as dividends. But there can be little doubt that the amount saved would have been less, perhaps very much less. The perennial inclination of minority stockholders to press for larger cash dividends and oppose proposals to reinvest earnings is evidence of this.

The activities of the insurance companies afford similar examples of corporate saving. The "level premium policy," the customary form in which life insurance is sold, results in the accumulation of immense funds in the hands of the insurance companies as reserves against their liabilities to the policy-holders. There is little reason to believe that the average policy-holder adjusts the amount of his contribution to the reserves of insurance companies by scrutinizing the rate of interest. Furthermore the practice of making the carrying of insurance compulsory, deducting the premium from the income of the policy-holder without his knowledge or consent, as for example in the case of legally enforced accident, sickness, unemployment, and other forms of social insurance, is spreading among industrial nations. This movement is tending to increase the proportion of society's investment fund whose accumulation bears no direct relation to the interest rate.

Of similar import is the practice of leaving estates in the hands of trustees or establishing trust funds for the benefit of the minor children or other relatives of wealthy men. In many such cases the trustees are bound by the terms of the agreement to reinvest a portion of the annual income regardless of the wishes of the beneficiary. Other examples could be cited. Loanable funds arising from such sources as these are added at any given time to those

¹ A. A. Berle and G. C. Means, *The Modern Corporation and Private Property*, 1933, p. 41.

resulting from voluntary choice of individuals and tend to increase the amount offered at various rates of interest.

The demand for loans. Turning now to the other side of the market, we note at the outset that the demand for loans comes from two distinct types of borrowing; *viz.*, *borrowing for production* and *borrowing for consumption*. The latter is for the sole purpose of increasing present command of consumable goods at the expense of future income. The first type arises among business men, whose borrowing is associated with productive enterprise. This type is by far the more important of the two and is the one which exercises the predominant influence upon the demand side of the loan market.

Borrowing for production: Purpose of the borrower. Unlike the spendthrift, the business man's loans are expected "to pay for themselves" and in addition to confer a gain upon the borrower. It is the purpose of the business man to exchange the borrowed purchasing power for certain forms of wealth and services to be actively employed in his enterprise — a new set of machines perhaps, or an additional building, or the paraphernalia of an advertising campaign. If we follow any one of these possible lines of expenditure to its conclusion, we discern that the loan has been employed by the enterpriser to divert a portion of the productive energy of society from the direct into the roundabout process of production. If additions to plant are constructed under the enterpriser's own management, a portion of the borrowed money is transferred to laborers in the form of wages on which to subsist while at work on a non-consumable product; another portion, expended in the purchase of materials, goes to reimburse someone else for a similar transfer of purchasing power to laborers. If these capital instruments are bought ready-made, the expenditure of the borrowed money gives support to those industries in which labor and capital are devoted to improving the preliminary stages of the process of production. Essentially similar is the result of a loan when it is used to buy intangible services rather than concrete articles of wealth. The enterpriser may have borrowed to pay the salaries of advertising agents, salesmen, or efficiency engineers, or the fees required to obtain a charter or franchise from the state,

or the expenses of a promoter who organizes a new industry or of the attorneys who draft articles of incorporation. All funds borrowed for productive purposes are ultimately resolved into advances to laborers.

The productivity of capital. The reader will recall that labor applied in the roundabout way is generally more productive of future income than labor applied to direct production. This roundabout process, to be sure, may not actually prove more productive in every case; labor may be wasted in overelaboration of the intermediate stages of production. But the possibility of net gain from the capitalistic process exists, and this possibility is increased by the dynamic character of our economic order. The borrowing of enterprisers rests on this fact of the productivity of capital.

But the enterpriser will ask something more of his borrowed capital than that it increase the physical product of his enterprise. He is not consciously concerned with the broad social aspect of the use of capital; what he is after is a money income for himself, to be derived from the sale of the products of his enterprise. He will borrow only if he expects to sell the increased product of his enterprise at prices high enough to enable him to repay with interest the money expended in the purchase or construction of his capital equipment.

Now the productivity of capital conforms to a law of diminishing marginal productivity, whether viewed from the standpoint of society or from that of the individual business man. In taking the social point of view, we must define the productiveness of capital as its power to increase the output of goods per unit of labor expended. If only a small quantity of capital is available for the use of society, it will be employed to equip men with the most essential labor-saving devices; for example the implements used to cultivate the soil, the wheeled vehicles, the tools of the craftsman, whose possession redeems modern man from the abject poverty of primitive races. As the quantity of capital increases, other things remaining constant, it will be applied to uses which grow less and less vital to the material welfare of mankind, until the last instalments of the capital supply of a rich society take the

form of insignificant improvements on a highly complex machine equipment, effective, no doubt, as labor-saving devices, but in no wise comparable in effectiveness with the first instalments of capital.

To the individual enterpriser all this is pertinent, and in addition he must recognize that the price of his product may fall as the quantity put on the market expands. Whatever the form of capital instrument or productive service to which he devotes his borrowed purchasing power, he will find each one subject to the law of diminishing productivity expressed in terms of money income. Moreover after he has devised that combination of productive elements which results in the greatest efficiency for the enterprise as a whole, to continue expanding the size of this balanced organization will eventually result in diminished money returns from successive increments of expenditure.

From the point of view of the business man, borrowed funds are thus subject to the general law of diminishing marginal utility which we have seen controls individual demand for all kinds of goods. The individual demand curve of the enterpriser-borrower conforms therefore to the shape of the normal demand curve, indicating that the amount that would be borrowed, at any given time, varies inversely with the rate of interest and is such that the marginal productivity of the capital thus obtained is equal to the rate of interest paid.¹

Borrowing for consumption. There are those whose present needs appear so urgent that they will seek to obtain the present means of satisfaction at the expense of their future incomes and will be willing to pay a premium for the exchange. Such borrowing is for consumption. Repayment of the loan and payment of interest must come, not from income derived from utilization of the loan funds, but out of the borrower's future income from other sources. There was a time, before the rise of modern capitalistic industry, when most borrowing was of this sort. Today such

¹ The importance attached in this discussion to the productivity of capital should not be allowed to mislead the reader into the notion that this constitutes an explanation of the existence of interest. As previously developed, it is time preference which, by limiting the supply of present goods available for loans, creates and perpetuates the phenomenon of interest.

borrowing is relatively insignificant, not because of any decrease in its absolute amount, but because of the preponderance of the other type of borrowing.

Consumption loans include the borrowing of those who resort to the pawn shops or buy goods on the instalment plan. Certain types of banks, for example the Morris plan banks, cater to such borrowing. Consumption borrowing is by no means limited to the very poor. Well-to-do persons, enjoying moderate present incomes but anticipating large incomes in the future, are inclined to borrow for present enjoyment. Anyone whose income is irregular may find it to his advantage to finance his personal expenditures by occasional short time loans. Most of the borrowing of political bodies is for purposes of consumption.

Consumption borrowing determined by time preference. Whereas the business man's demand for loans is related to the anticipated productivity of capital, the corresponding factor in the case of the borrower for consumption is his time preference. A high rate of time preference leads to eagerness to borrow. The various conditions, of individual character or of external circumstances, which lead anyone to have a high rate of time preference may therefore be considered as influences impelling him to borrow.

So long as the borrower's rate of time preference is higher than the rate of interest which he has to pay for loans there is gain to him from borrowing. But borrowing for consumption has the effect of enriching the present and impoverishing the future. In accordance with the general law of diminishing marginal utility, each unit of borrowing adds a progressively smaller quantum of satisfaction to the present income and entails a progressively larger future sacrifice. The marginal utility of the present gain decreases, while the marginal utility of the future income sacrificed increases. It follows that the individual's rate of time preference, based as it is on the comparison of these two marginal utilities, tends to decline as he borrows, until eventually it becomes equal to the rate of interest. Thereafter borrowing ceases to be advantageous. Hence the amount an individual is willing to borrow at any given time tends to vary inversely with the rate of interest. His demand curve therefore is of the same general shape as that of the borrower for pro-

duction, and at equilibrium his marginal rate of time preference is equal to the rate of interest he must pay.

Borrowing of governments. The borrowing of political bodies is generally similar to the borrowing of individuals for consumption, although some government loans are for purposes of production. In the typical case, the government decides on grounds of social or political expediency to embark on a line of action from which no future financial returns can be expected. The deficit resulting from this expenditure is then covered by the sale of government bonds or other evidences of indebtedness. There is also much short term borrowing in anticipation of receipts from taxes whose income is unevenly distributed throughout the year. The government, like the private individual who borrows to obtain satisfactions in excess of his present means of payment, is in such cases merely anticipating future revenues in order to finance large present expenditures.

But though these government loans are similar in nature to the consumption loans of private persons, their amount is not determined by the same forces which control the action of the individual with reference to the loan market. The amount of loan funds demanded by political bodies does not vary in any predictable way with the interest rate. A high rate of interest will sometimes cause the postponement of some government enterprise for which borrowing would be necessary; it is more likely to affect the tenure of the loan; *i.e.*, whether it runs for a long or a short term. In time of war, the statesmen in charge of the nation's destinies simply borrow all that is needed to bridge the gap between tax receipts and war expenditures, or all that the lenders can be induced to provide for this purpose, paying the lowest rate of interest practicable.

Total demand for loans. The demand of enterpriser-borrowers and that of individuals who desire to increase their present enjoyable incomes may be thrown into the form of schedules and combined to form a total demand. The general rule governing both of these demands is that the amount of the loan varies inversely with the interest rate. Government borrowing is not easily reduced to a demand schedule because non-economic motives weigh heavily in all political activity. The government comes into the market as

occasion requires, with a predetermined demand for loans, and offers the lowest possible rate of interest. Though there are times when the necessitous borrowing of governments is controlling, as a rule the demand of enterprisers so vastly outweighs all other types of borrowing that it governs the market. Borrowers for consumption and governments in ordinary peace times must adjust their demands to a loan market dominated by business men whose borrowing is based on the productivity of capital. Strictly speaking however all three types of demand must be added together to give a measure of the forces operating at any given time on the demand side of the market.

Determination of the market rate of interest. As already noted, the rate of interest is a price, the price of waiting, determined

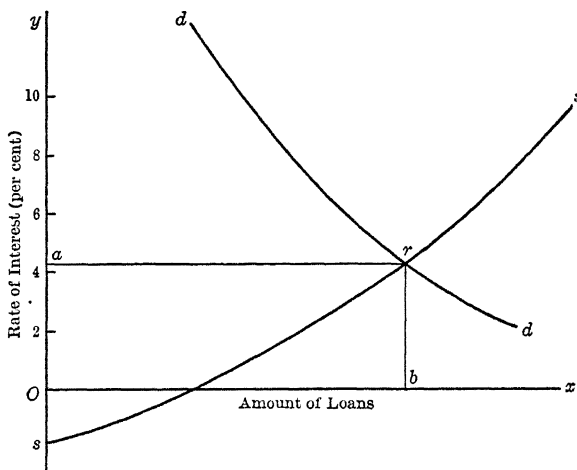


FIG. 35

in the loan market. Such peculiarities as attach to this particular price have now been disclosed by our study of the demand and the supply of loans. We may conclude this part of our investigation by stating that in any given market at any given time the rate of interest and the amount of loans is determined by demand and supply, as illustrated in Figure 35, in harmony with the general laws of price as set forth especially in Chapter XIII of this book. The demand and supply curves have the normal slopes; the part of

the supply curve below the base line illustrates the fact that some lenders would, if they had to, be willing to lend without interest and even to pay something for the safekeeping of their savings.

Adjustment of the individual to the loan market. The rate of interest, like any other price, is the effect of the activities of a multitude of individuals demanding or supplying loan funds. But again like other prices, this rate is also one of the causes of the amount offered or taken by any given person. The individual looks upon the existing rate of interest (as he does upon any other price where pure competition prevails) as fixed by forces beyond his control; he is concerned with adjusting his activities so as to make the best of conditions as he finds them. In theory, certain principles may be laid down which control these individual adjustments.

(1) *The individual lender.* A man will lend when his rate of time preference is lower than the rate of interest. He will continue to lend until his marginal rate of time preference is equal to the interest rate. Theoretically he gains from all units of his loan except the last and, by lending up to this point, makes his gain a maximum. In a perfect loan market, the result of the activities of the individuals will be to give them marginal rates of time preference equal to each other. As noted in the preceding discussion some supplies of loan funds are drawn from individuals who save without reference to the interest rate and from others who have no free choice in the matter. This portion of the total supply is not controlled by this principle.

(2) *The enterpriser-borrower.* This person borrows when the anticipated marginal productivity of capital under his management exceeds the interest rate, and he continues to borrow until this discrepancy is removed. Since it is the income from the marginal unit which equates with the interest rate, there is a gain to be derived from the use of borrowed funds for productive purposes. This surplus is not all retained by the enterpriser-borrower. It represents that part of the income of the enterprise which is not absorbed by the interest cost, and it is distributed among the other productive factors which comprise the business unit. The tendency for each enterpriser to borrow until the marginal income from

the loan equals the interest rate results, in a perfect loan market, in distributing the capital supply of society among the different business units and the different forms of capital instruments so that its marginal productivity is everywhere the same. This tendency works toward the end that society receives the maximum product from a given capital equipment.

(3) *Consumption loans.* An individual will borrow for consumption when his time preference rate is higher than the rate of interest, and he will continue to borrow until this discrepancy is removed. Like the buyer of goods, he gains utility from all units of his loan except the last one. In a perfect loan market the tendency is for the rates of time preference of all such borrowers to reach equality. As stated above however the borrowing of governments is not generally controlled by this principle.

(4) *The alternative of lending or borrowing.* Many people are in a position to be either borrowers or lenders, the choice depending upon the relation of their own rates of time preference to the market rate of interest. Let us suppose that a certain person's rate of time preference is six per cent. If he finds that he can borrow at five per cent, it will be to his advantage to do so. But if the market rate of interest is seven per cent, he will have no inducement to borrow. On the contrary, if he is the possessor of any considerable capital, he will find it to his advantage to lend some part of his present resources, up to the point at which his own rate of time preference becomes seven per cent. In particular the business man who is himself an owner of capital will enlarge his enterprise with borrowed funds if the market rate of interest is lower than the marginal productivity of his enterprise, whereas he will contract his operations and lend his free capital to others when the opposite relation holds. The possibility of shifting from one side of the market to the other contributes much to the sensitiveness of the loan market. A practical demonstration of the importance of this factor was given by the New York money market between the years 1924-1929. So many large corporations, formerly borrowers, were impelled by the high interest rates to become lenders that loan funds from these sources became the controlling factor on the supply side of the call loan market.

The rate of interest and the quantity of money: Traditional fallacy. We may now dispose of a stubborn fallacy, which persists despite repeated disproofs. Each generation brings to the front a group of reformers who, either themselves debtors or impelled by a compassion for the debtor class, propose to abolish interest or to lower the rate of interest by increasing the quantity of currency. Their program is based on the belief that the rate of interest varies inversely with the quantity of money, this belief resting in its turn on the assumption that borrowing constitutes a demand for money and that therefore the price which the borrowers pay (*i.e.*, the rate of interest) must be controlled by the relation of this demand to the supply of money. This assumption would not be far from the truth if the term "supply" were used in a sense coördinate with the term "borrower's demand"; if, that is to say, it referred to the supply of money available for loans. But we have seen that the borrower does not want money itself, but money's worth. Since the effect of an increase in the quantity of money is to reduce its purchasing power, a larger amount will be required to accomplish the purpose of the borrower. It follows that, as the price level rises in consequence of an increasing supply of money, the demand for loans measured in dollars and cents also rises, thus tending to offset the effect of the increased supply.

Bank reserves. Experience may seem to bear out the opinion that the short-term interest rate fluctuates with the amount of money in the country. The transactions of borrowers are carried on to a large extent with the banks, and their knowledge of the loan market is formed out of their experience with these lending institutions. From this experience they learn that when the quantity of money in the bank's vaults rises or falls in proportion to loans carried by the bank, the interest rate on loans tends to move in the opposite direction. The banker will explain the low rate of interest which accompanies a relatively large cash reserve by the statement: "money is plentiful," and the high rate which accompanies the relatively small cash reserve by the statement: "money is scarce." But here we have to do, not with the amount of money in the country, but with the amount in the banks and with the motives which induce people to deposit their money as an alternative to spending it. If the bank

reserves increase only because of an increase in the total stock of money — that is, if the proportion between the amount of money in the banks and that in circulation remains the same — the greater lending power of the banks will be offset by the increased amount of money required by the borrowers to conduct their transactions on the higher price level. If however the bank reserves increase faster than the amount of money in the country, this is an indication that the people, for one reason or another, are increasing the proportion of money available for loans. In this case there will be a tendency for the interest rate to fall, but obviously the cause of the fall is to be sought in the motives which have induced the people to save more and spend less, not in the amount of money in the country at large. Nor can it be shown that an increase in the total quantity of money will enable the people to save proportionately more than formerly when account is taken of the rise of the price level which money inflation causes.

Transition effects. The conclusion we have reached has to do with the ultimate effects of changes in the quantity of money. It still remains to inquire how the interest rate is affected by the transition phenomena of the period during which the quantity of money is increasing or decreasing. We know that when prices are rising under the influence of currency inflation, business men strive to take advantage of the favorable market by expanding their enterprises through increased borrowing. If under these conditions the new stocks of currency are so distributed that the bank reserves temporarily fail to maintain their ratio to the amount of money in the country, the need for loans will outstrip the quantity of money available for loans and the interest rate will rise. On the other hand, when prices are falling because of a contraction of the currency, money may find its way into the banks in unusual amounts, due to the tendency of the people to restrict their expenditure during periods of depression. At the same time, the general feeling of business uncertainty will tend to reduce the demand for loans; hence the interest rate may fall when the quantity of currency is decreasing. The effect on the interest rate, in other words, will be the opposite of that expected by those who believe that the rate of interest varies inversely with the quantity of money.

We should note also that the effect of a rising price level is to reduce the real worth to the lender of the principal of loans which he made prior to the rise of prices. A loan of a thousand dollars loses one-half of its purchasing power if the price level doubles before the loan is repaid. Rising prices therefore penalize lenders in the case of all pre-existing loan contracts and benefit the borrowers to an equivalent extent. Conversely a falling price level impoverishes borrowers and enriches lenders. It is to be expected that borrowers and lenders will take account of these effects of a fluctuating price level and attempt to guard against them by raising or lowering the interest rate. During a period of rising prices the interest rate tends to rise as an offset to the depreciation of the principal of the loan; during a period of falling prices it tends to fall as an offset to the appreciation of the principal. This tendency coöperates with the one just discussed to cause a rising interest rate when the supply of currency is increasing, and *vice versa*. These, to repeat, are transition phenomena. They disappear when the price level is again stabilized.

Lack of uniformity in interest rates. Thus far we have spoken of the interest rate as if there were a single uniform price applying to all types of transactions in the loan market. But when we turn to the business world for illustration of our theory we are impressed at once by the absence of any such uniform price. Different rates of interest appear in the market at the same time, each rate applying to a particular type of loan transaction. For example the following schedule of rates was published on three different dates for transactions in the New York money market:

Type of loan	Per cent		
	JULY 26, 1929	FEB. 16, 1935	JULY 15, 1938
Call loans.....	10	1	1
60-90 day loans.....	8	$\frac{3}{4}$ and 1	$1\frac{1}{2}$
Commercial paper			
Best names.....	6	$\frac{3}{4}$	1
Other names.....	6	$1\frac{1}{2}$	$1\frac{1}{4}$

Some of the variations appearing in the first item of this table are due to conditions peculiar to the banking business. As we know, a large part of the banker's liability consists in deposits payable on demand against which he must maintain assets in liquid form.

Bankers in a money centre like New York receive on deposit large amounts of cash funds from banks in inland cities subject to continuous withdrawal and replenishment. For obvious reasons it is not profitable for these bankers to hold such deposits in the form of idle cash. The call loan is the best form of investment for funds of this kind; it is resolvable into cash on demand and secured against loss by a deposit of collateral. For this reason the call loan in a large money centre usually commands a peculiarly low rate of interest. This rate also fluctuates sharply in response to variations in the amount of money seeking this particular form of investment. The student will realize that the interest rates listed in the second and third columns of the table indicate an abnormal condition of the money market. Prolonged business depression had caused the banks to carry an unusually large portion of their assets in cash. This abnormal supply of short-term loan funds coupled with the decline of demand in the business world resulted on the dates given in extremely low rates of interest in the New York money market.

True and nominal interest rates. On loan transactions not subject to these exigencies of the banking business, variations in the interest rate are often more apparent than real. We must view interest as a payment for a single kind of service; *i.e.*, the service of postponing income by a voluntary exchange of present goods for future. If the man who renders this service performs at the same time a number of other functions and receives a single payment as a recompense for all of them, only a part of the payment can strictly speaking be called interest; the remaining part is composed of as many different elements as there are different functions to be rewarded by the payment. Now it happens that the term "interest" is commonly used in the business world to define just such a composite payment, of which only one element is true interest.

The rate of interest in ordinary loan transactions contains, in addition to interest proper, an element which should be considered insurance against risk. When dealing with true interest we must assume that lenders are assured of their future return. If in addition to sacrificing present income they must bear risk of loss, their promised reward must be increased by a payment suf-

ficient to induce them to take the risk. The differences in the rates of interest earned for example on different types of bonds may be attributed chiefly to this element of risk. For our purpose it may be assumed that through all these different types of investment runs a common true interest rate, which is augmented in each case in proportion to the risk involved.

The capitalist who undertakes to manage his own investments must perform other duties in addition to postponing the enjoyment of his income. He must expend effort in searching out profitable lines of investment, collecting information relating to the industries they represent, following the trend of market conditions in order to safeguard his interests. In certain types of investment, as for example the ownership of rented houses, office buildings, etc., he must also undertake to collect the interest periodically and to care for the capital. In all such cases true accounting would segregate a portion of the return which is loosely denominated interest and give to that portion the name wages of management. The fact that a separate group of men, investment brokers, trustees, etc., has developed to relieve the lender of some of these duties, in consideration of a stated remuneration, shows that the duties are not inevitably combined with the services of the lender. When the investor assumes these tasks himself, it is proper to consider him entitled, in addition to interest, to the same remuneration that he would have been compelled to pay another to perform such services in his behalf.

Wages of management account for an even greater part of apparent interest when the business man uses his capital in his own business and, in his cost accounting, either neglects to allow himself a salary or fixes his salary at a figure below his real worth as a manager.

Yet another element commonly confused with interest is depreciation. When an investment takes the form of capital instruments which are wearing out, a portion of the capital value is returned with the interest each year, thus making it appear larger than it really is. If this continues without the knowledge of the investor, he will find in time that his capital no longer exists; it has been consumed in the process of returning an annual income. The

prudent business man will anticipate this result and provide against it by setting aside a certain part of the income in a depreciation account, viewing only the remainder as true interest.

In the study of the true rate of interest then it is essential to eliminate the other payments which are commonly confused with it. An investment whose nominal and true rates of interest are identical is one in which the investor bears no risk, exercises no managerial powers, and receives no depreciation in connection with the income. Very probably, as a practical matter, no investment can be found which measures up exactly to this ideal. But it is approximated in the case of the long-term bonds of a stable government, whose return may be accepted as a fair exemplar of true interest.

Changes in the rate of interest. If we examine the trend of the interest rate over long-run periods, we are impressed by its relative stability. When in the seventeenth century the interest rate in Holland fell to three per cent, it stood at a point only slightly lower than during the flourishing days of the Roman Empire. In the following century the rate rose slightly, only to fall during the nineteenth century to the same low point. The five decades preceding the World War witnessed some rise in the rate. With this record may be compared the yield of high-grade securities at the end of the nineteenth century, as shown by a recent study of the average interest return on government bonds over a 21-year period centering in 1896.¹

<i>Bond</i>	<i>Average Yield</i>
British Consols	2.77
U. S. Bonds	2.22 *
French Rentes	3.16
German Bonds	3.46
U. S. Rails	3.51

* It should be noted that this rate for United States bonds is abnormally low, owing to the valuable privilege carried by these bonds of serving as security for national bank note issue.

But this stability over long ranges of time contrasts with fairly sharp fluctuations in the rate in short-run periods. The rate for current loans is a market price which tends to respond to momentary changes in the factors of demand and supply. There is also

¹ G. F. Warren and F. A. Pearson, *Prices*, 1933, p. 278.

a tendency for the rate to rise in the wake of rising prices during the "prosperity" phase of the business cycle and to fall to a low point during the subsequent reaction. But these responses of the interest rate are at best sluggish and retarded in action, a fact which has unfortunate economic consequences. Of course the rates of pre-existing loans, being determined by legal contract, do not change at all in response to fluctuations of the business cycle or change in a disjointed manner at the end of the respective loan contracts. Moreover the rates on current loans are so influenced by customary business practice and by imperfect competition in the money market that they too fail to move promptly and smoothly in harmony with changing business conditions. If the price of loans were to rise promptly during the early stages of rising prices, it would act as a brake on unhealthy credit expansion and thus tend to prevent a subsequent crisis, but because of these market frictions it fails to keep pace with the situation. Similar lagging of the interest rate may be observed during the early stages of the business depression which follows a crisis. The general collapse of prices at such a time reduces the value of business assets and current business income, while the sluggish behavior of the interest rate prevents a corresponding reduction of business indebtedness and business costs. This tends to undermine the solvency of enterprises which are in debt as well as to make it more difficult to place current business transactions on a profitable basis. If the depression is prolonged, the rate may fall to abnormally low levels as at the present time (1939), because the continued lack of demand from the business world allows the supply of loanable funds to outstrip the demand for them.

The capitalist class. There is no doubt that the rôle played by capital in the productive system of society must grow in importance with the passing years. Our civilization is founded on the use of great masses of capital; progress in human well-being — nay, the mere support of life for an increasing population — requires that we set aside each year an enormous amount of present income for the repair and replacement of existing capital instruments and the addition of new ones. The industrial mode of production calls for two distinct kinds of service: working and waiting. Neither of

these services by itself can maintain the flow of consumable income at a volume sufficient to satisfy the wants of men on the present plane of living. Man's labor of hand or brain, unless reinforced by his willingness to wait and to endure the costs of waiting, must be applied to direct production without the help of those product-multiplying aids which we call capital instruments.

Any commonplace incident of modern productive life will illustrate the indispensability of the service of waiting. Let us suppose that a group of people who possessed no resource other than their labor and their intelligence undertook to produce the houses required to alleviate the overcrowded condition of the slum districts of our great cities. Their first step would be to appropriate from the resources of nature the necessary supplies of lumber, ore, stone, cement, and other building materials. Each of these must then be fashioned by a time-consuming process into forms suitable for the purpose in view. More time must elapse while they are transported to the places where the want for shelter is felt. Then must begin the work of construction looking toward the creation of consumable income. At no stage in this long process would the laborers be able to subsist on the current product of their labor. If we relied on labor alone, the undertaking must be abandoned at the very start, and no houses would ever be built. When therefore some member of society who is not in the group of workers comes forward with an offer to "finance" the undertaking; *i.e.*, to transfer a part of his present income to the workers in exchange for their non-consumable products, it is obvious that his service of waiting is as indispensable to the production of this necessity as the services of the workers. When the productive process is finished, this individual will own the capital instruments which have been produced and will draw from their use by others a net income as his payment for waiting. In other words, he will be a typical "capitalist," sharing in the social income by right of ownership.

As long as society relies on voluntary saving for the provision of capital, the numbers of those who draw interest incomes must increase, and the size of the incomes of this type which accrue to the larger capitalists must expand as the productive process grows more capitalistic. Indeed it is natural and appears to be the fact

that the proportion of the total national income accruing to capital owners tends to increase as improvements are made in the technological efficiency of industry.

Although, as we have seen, the service of waiting is an indispensable productive function, this service is not rendered at the time the interest income is received but has been performed in the past, when the saving was done. Indeed in many cases the service of waiting is not even performed by the same man who receives the interest but by some predecessor from whom he has inherited his wealth. One consequence therefore of the creation of a distinct capitalist class is the existence of a group of people who appear to enjoy the right of sharing in the income fund of society while making no concurrent contribution to that fund.

Opposition to interest as a private income. The spectacle of the receipt of incomes by individuals, some of whom appear to be idlers, affronts the sense of justice of many a man who does not reason clearly with regard to the capitalist's function in society. Other grievances associated with the existence of interest as a private income add to his bitterness and give impulse to his attack upon the institution. By inheritance ownership of capital falls to men and women who live in opulence as drones in the midst of a busy world. Powers other than the mere right to receive interest are gathered into the hands of the capitalists, power of command over the lives and welfare of property-less men and women, power to influence for good or ill the functioning of society's industrial structure, preponderance of power in the political world, all of which may give the capitalist a sinister aspect in the view of the common man. Reaction against the institution of interest as a private income leads to organized movements which propose a thoroughgoing change in the present economic structure of society.

Regulation of interest: Early experience. Later we shall examine in some detail the reform movements associated with the term, socialism. However some aspects of the attempts to regulate the interest income have so direct a bearing upon the subject in hand that we must take some account of them here. In part the modern attack upon interest is an inheritance from an earlier day and a simpler stage of societal development. Throughout recorded

history there have been individuals and communities which have condemned the receipt of interest by lenders as an immoral and indefensible practice and have taken steps to abolish or strictly to regulate it. Aristotle decried the taking of interest on the ground that money was "barren" and could have no legitimate "offspring." The Mosaic Law prohibited the payment of interest between Jews. The leaders of the early Christian Church carried over this condemnation and applied it to their own communities. By the laws of Rome interest-taking between Roman citizens was similarly forbidden. Perhaps the best known example of this attitude is supplied by the Church during the Middle Ages. For over a thousand years, indeed down to the thirteenth century, the Church Fathers carried on an unceasing attack upon interest. By the Canon Law its receipt was strictly forbidden to Christians on pain of spiritual punishment in this world and the next, a prohibition which had the incidental effect of throwing the business of money lending into the hands of the Jews, who were considered in any case to stand outside the pale.

These laws became obsolete and gradually disappeared in all modern nations. But vestiges of them remain in the form of "usury" laws, and the point of view which they represent comes to the front in modern times, not only as a phase of the program of the socialists, but also in the form of sporadic political movements among more conservative men. During the earlier period the term "usury" defined the entire amount of interest received by the lender; all of it was condemned as an immoral income and all prohibited. Today we divide the interest payments into two parts and employ the word usury as a term of condemnation applied to the excess of the interest rate above an amount vaguely fixed by changing ideas of justice and fair play. The usury laws of the American states, which prohibit the taking of interest above a fixed maximum on certain types of loan transactions, are of this character.

There was some ground for moral condemnation of the interest income in early stages of social development. Almost all loans in that day were loans for consumption. The spendthrift borrowed to enlarge his supply of consumable goods; the necessitous man, the

widow, the victim of misfortune were compelled to borrow to eke out a day-to-day existence. The king or the lord of the manor borrowed to cover the expenses of a war or crusade or to heap favors upon the court favorite of the day. Such loans were indeed "barren," to adopt the expression of the moralists of the time. As long as they remained the typical form of loan transaction, it was difficult either to explain on economic grounds why interest existed or to justify its existence on grounds of morals. Certainly such a borrower was not enabled to bear the charge which interest laid against his future by creating income through his use of the borrowed funds, as does the enterpriser-borrower today. And in the common run of cases the position of the creditor gave him the appearance of exploiting the necessities of his brother man. The usury laws of the time were passed in defense of the weaker man, the debtor, to protect him from being surrendered by misfortune into the hand of the voracious money lender.

The modern field for regulation. Where this type of loan has carried over into modern society we continue, with general approval, to regulate the interest rate out of concern for the borrower. In the industrial cities of all countries there exists a class of lawful pirates who have acquired the illuminating title of "loan sharks." They flourish among the timid and ignorant and most necessitous sections of the wage-earning class, taking advantage of the proverbially high time preference of this class to involve them in bonds of indebtedness, exacting exorbitant rates of interest for petty loans advanced usually on the security of chattel mortgages. Familiar to all city dwellers is the pawnshop of such districts, another type of lending agency which caters to an occasional but highly urgent want of present income among people who, for one reason or another, are excluded from the general loan market. In many countries whole sections of the people tend to fall into the clutches of an inhuman creditor through their borrowing for improvident purposes. Though interest is no longer prohibited on loans for consumption, all modern countries make some provision for the protection of the most necessitous borrowers from the worst consequences of their thriftlessness. France has established public pawnshops, where small loans may be had at reasonable rates of

interest. England has opened quasi-public banks in the rural districts of India and Egypt, which underbid the village Shylock both in terms of interest charge and in facilities for discharging the indebtedness. In America the states regulate the pawnshops with laws fixing a maximum interest rate and requiring accessible records of all transactions, while public-spirited citizens have set up a system of loan offices designed to break the hold upon his unfortunate debtor of the grasping lender on chattel mortgages.

Futile and injurious interference with the interest rate. But the typical debtor of today is not the timid victim of misfortune whom imagination pictures as struggling in the grip of a brutal money lender. On the contrary he is the most able, most self-reliant of us all, the business enterpriser. As modern society emerged from the Middle Ages, the capitalistic system of production began more and more to provide the outlet for the accumulated funds of the savers. Borrowing for production in the normal course of events places the borrower in position to meet the interest charge without trenching on the income which otherwise would accrue to him. However despite the fact that the day has passed when spendthrift loans comprised the major part of the business of the loan market, the antipathy toward interest, especially a high rate of interest, remains and finds expression in laws intended to restrict the rise of the rate in the financial world. Admitting that exceptional cases arise which require exceptional treatment, it should be evident that such laws are in general both futile and injurious. They are futile because, when conditions arise which enable the enterpriser-borrower to pay high rates of interest and therefore create in him a disposition to do so if necessary, the purpose of usury laws will be defeated by connivance of both borrower and lender. They are injurious, both from the standpoint of the individual borrower and from that of society as a whole. If any branch of business or any geographical locality is handicapped by lack of capital, and this is invariably the case when complaints are heard about the high rate of interest, it is obviously to the advantage of all enterprisers to facilitate an increase in the supply of loan funds. The basic cause of the condition cannot be removed by arbitrarily relieving present debtors of a part of their interest costs. What is needed is renewed

inducement to saving and investment. Usury laws, by depriving savers of a part of their inducement to save and by injecting into the loan market the incalculable element of political interference, tend to perpetuate the evil which lies at the root of the protest, a paucity of loan funds. Not only the individual who is attempting to carry on productive enterprise under the handicap of small capital supply, but the entire community which shares in the benefit of the capitalistic process, is benefited by allowing the interest rate to fluctuate freely in accordance with the forces of demand and supply.

It will be understood that these statements are made with reference to the type of society with which we are familiar — one based on private property and individual economic enterprise. A social order fundamentally different from this, such as is proposed by the socialists, would place the problems of capital and interest in an entirely different setting.

XVIII

THE GENERAL LAW OF WAGES

The rewards of human exertion. The vast majority of mankind must derive their income from their own exertion; that is, from labor. The term *labor*, as generally used in economics, has a broad meaning, including specifically the exertions of the members of these three groups: (1) those who sell their labor, whether manual or mental, to employers in return for wages or salaries, to which class the term labor in a narrow sense is sometimes confined, (2) independent professional men and women, such as lawyers, physicians, artists, and technical experts, who render services to their clients in return for fees and professional earnings, and (3) business men, enterprisers, entrepreneurs, who own and direct industrial enterprises and receive, in addition to interest and profits, rewards for active managerial functions, known as wages of management.

While it is customary to distinguish wages, salaries, fees, and professional earnings, it will be found convenient for economic analysis to lump these forms of income under the single general term *wages*, which, in that sense, may be defined as *remuneration received from other persons in exchange for labor*. This includes the incomes of all those in the first two labor classes, as distinguished from the incomes of the third class. The investigation of the causes which determine wages will occupy us during the next two chapters, following which we shall give our attention to the subject of profits.

The problem of the rate of wages. Since wages result from the sale of labor, the rate of wages is a species of price, governed by the forces of demand and supply in the labor market. But a study of these controlling forces presents peculiar difficulties, which make an explanation of the wage rate one of the difficult problems of economics. We search in vain through the business world for a uniform rate of earnings for different workers similar to the rate of interest which, with relatively small and explicable variations, determines the incomes of different capitalists. The great body of the

working people is divided into a bewildering number of groups and sub-groups, each enjoying a different rate of income, and within these separate groups the earnings of individuals will be found to vary in such fashion that any attempt to formulate a general law governing the rate of wages appears hopeless at the outset.

In view of these complexities, the first step must be to simplify the problem. This can be done first by limiting attention at the start to some group within the great class of wage earners whose members are fairly homogeneous in quality; that is to say, a group so alike in function and capabilities that one man can be substituted for another without material loss to the industry. In the second place, we shall assume a condition of free competition and a perfect labor market and develop the theoretical principles which would operate under such confessedly artificial conditions. Under these assumptions, there will tend to be a uniform wage rate for the group, determined by demand and supply. The qualifications required to adjust this simplified picture to the complex facts of the industrial world can be made after we have analyzed the forces which tend toward a uniform wage rate within this section of the labor market.

The demand for labor: Why employers hire labor. The wage rate for any given labor group is dependent upon the nature of the demand for labor, the peculiar characteristics of labor supply, and the joint action of demand and supply in the employment market. Wages are paid by profit-seeking employers, who voluntarily undertake to give the workers a money income in exchange for their services. Each employer assumes responsibility for paying wages for the same reason which impels him to burden his future income with interest charges when borrowing capital. That is to say, he expects the workers to create the income out of which wages will be paid, just as, when borrowing capital, he expects to increase his income by an amount at least equal to the interest charge. In both cases he has but one source from which to recover his costs; *i.e.*, through the sale of the product of capital and labor. The employer's demand for labor then arises out of his expectation of money for labor's product, in other words, out of the anticipated monetary productivity of labor.

Marginal productivity of labor. In any homogeneous labor group such as we are considering it is not the total but the marginal product of the group which governs the employers' demand for labor. Employers are not confronted with the choice of hiring the entire labor force as a single unit or hiring none at all. They employ laborers as individuals and are concerned with the specific contribution of the individual worker to the income of the enterprise. The amount of this contribution is measured for each member of a homogeneous labor group by the *marginal productivity* of the group; that is, *the extent to which the total product per laborer changes with a small change in the number of workers*.

The productivity of labor conforms to a law of diminishing marginal productivity. There are two reasons for this. As appeared in our study of interest, when the quantity of capital increases while the coöperating factors remain fixed in quantity, the physical product of the combination tends to fall as measured in proportion to the number of units of the increasing factor. So it is with labor. The physical product per unit of labor will eventually decline as the number of laborers is increased. In the second place, beyond certain limits increasing quantities of output must be sold at lower prices. It follows that the contribution which each unit of labor makes to the income of a business enterprise tends to fall as the employer expands his labor force, other factors remaining constant. This is in conformity with a principle with which we have become familiar in connection with our study of the other functional shares in distribution. Applied to labor, it means that the employer cannot go on expanding his labor force, land and capital being constant, without noting a decline in the amount of income attributable to one unit of labor.

Under these conditions the employer, confronted with a given rate of wages, will add to his labor force until the contribution of the marginal man in each group to the income of the enterprise is no greater than his wages. In simple language this means that the employer will continue to hire men as long as he gains by so doing, and he gains as long as the marginal man's product is in excess of the wage rate. Evidently the lower the wage rate the greater the number that can be employed before this limit is reached; hence

the number of laborers demanded by a given employer or by all employers under any given set of circumstances varies inversely with the rate of wages.

Productivity discounted. One important technical feature remains to be considered. The workers will want their pay in advance of their employer's sale of their product. Generally at the end of each week the laborers will ask to be paid for the services performed up to that time. But in the majority of cases the products upon which they have been working are not in a completed form and ready for the market at the time the wages are paid. This phase of the matter was touched in our study of interest, where the capitalistic process was described as a succession of advances to workers. But the employer will not thus make advances to laborers unless he is paid for his waiting; that is, unless he receives interest upon the advance payments. The wages which the employer can afford to pay the marginal worker therefore will fall short of the value of his contribution to the product of the business by the amount of the interest charge. Thus the employer is compelled to discount the marginal product of labor in deciding what wage he can offer.

The supply of labor. Other things being equal, the quantity of any good which will be offered for sale varies directly with the price. In general this is true of the supply of labor within any single occupational group and within the wage class as a whole; but it should be noted that the circumstances back of the supply of labor are in many respects peculiar. The amount of labor available at any given time is a comparatively fixed quantity, not susceptible to great variation in relation to the rate of wages. In other words, the total supply of labor is comparatively inelastic. Even if we consider a single homogeneous labor group, the inelasticity of supply is only slightly modified, since, as we have already seen, laborers are not easily drawn from one group into another, even by the inducement of considerably higher wages. Moreover changes in the total number of laborers take place slowly, and there is only in a remote sense such a thing as a cost of production of labor. Therefore though the supply of labor is similar to the supply of things in general, in that the quantity offered varies directly with

the price, nevertheless we must conclude that in the case of labor this variation is only moderate.

The general law of wages. These forces of demand and supply would, if unimpeded, determine a market rate of wages in a manner similar to the determination of other prices in a competitive market. The demand for each type of labor is a schedule of rates representing the discounted marginal utilities of labor groups of differing sizes. Because of its inelasticity, the supply of labor is at any given time virtually a fixed number of workers in each homogeneous labor group. The conditions of demand will fix the highest rate at which this number of workers can be absorbed in employment under the free play of economic forces. The general law of money wages may be stated as follows: *The normal rate of money wages for any uniform group of workers is equal to the discounted marginal productivity of that group.*

This is the basic law of wages. It holds true on the assumption that each employer must pay the prevailing rate of wages as determined by the general forces of demand and supply. Conditions peculiar to a given enterprise or locality may sometimes enable an employer to obtain a certain number of workers for less than this market rate. He may then find it more profitable to limit his work force to the number available at this lower rate than to push his demand for labor further. In this case the enterpriser in question will receive from his labor force a marginal product in excess of that prevailing in the general labor market. Other exceptional cases may be imagined or discovered to exist in the world of practical affairs; they are obviously due to imperfect competition in the labor market. Under perfectly competitive conditions the principle stated above would govern the reaction of each employer in the labor market.

But although the individual employer assumes that the wage rate is fixed by market forces extraneous to himself, he plays a part in causing this rate to be what it is. Were he to discontinue his enterprise and withdraw from the market, the workers whom he released would have to seek employment elsewhere, thereby (in theory) lowering the marginal productivity in other enterprises and reducing the wage rate. On the other hand, if this employer is

enabled by some development in his business to increase his labor force without reducing their marginal worth, his increased demand for labor can be satisfied only by diverting wage earners from other industries, thus raising marginal productivity in general and increasing the wage rate.

In strict theory no employer can obtain so large a supply of any type of labor that its marginal productivity in his enterprise falls below its marginal productivity in the enterprises of competing employers. If this should happen, a second employer would be able to offer higher money wages than the first and would draw labor away from him. When competition on both sides of the labor market has worked out a stable distribution of the labor supply, some employers will have more of this type of labor, some less; but in theory the marginal man in the working force of each employer will contribute substantially the same amount to the income of his enterprise.

Practical qualifications: Theoretical precision lacking. This is the theoretical view. But our general law of wages is obviously an abstract statement of a tendency, rather than an explanation of all the diversities which may be observed in the business world. We must now add qualifications in order to take account of some of the more important of these practical considerations. It is obvious, in the first place, that the appearance of precision given in our illustration to the employers' calculation of the worker's marginal productivity is artificial. In the complexities of modern industry with its elaborate division of labor no one can say exactly what one worker contributes to the corporate income. At best these calculations are intelligent guesses based on experiment and experience.

Interdependence of different types of labor. In the second place, it never works out that the different enterprisers enlarge their several labor forces to the point where the marginal productivities of similar grades of labor are precisely equal. The number of workers of a certain type which can be employed profitably in a given industry is affected, not only by the marginal productivity of these workers, but also by the necessity of maintaining an effective proportion between the different types of labor in the enterprise. Complex division of labor creates a situation in which no

single type of labor can be employed without necessitating the employment of other types. For example an industry employing one bookkeeper might not have need for the services of another at any rate of wages, so long as its present scale of operations remained unchanged. To make it possible for a second bookkeeper to produce anything might require the employment of a dozen machinists, a foreman, a shipping clerk, and two or three other types of labor, none of which could be employed profitably at their prevailing wage rates. In a certain large factory producing tennis shoes, one inspector is employed in conjunction with twenty-five workers in other lines; one cutter works with half a dozen stitchers, etc.

These examples illustrate the fact that the demand of any employer for a given type of labor is limited by his force of coöperating types, which are limited in turn by the wage rates of these other labor groups. He tends to increase the number of workers in each group up to the point where their marginal productivity equates with the wage rate, but he may stop short of this point because of his inability profitably to expand the scale of his industry. Obviously the small business is much more affected by this consideration than is the large one, which, with adequate supplies of capital, can take more complete advantage of the gains inherent in complex division of labor. When however attention is fixed on a single occupational group, it is probably true that in no individual enterprise could there be found that precise equivalence of marginal productivity with the wage rate which our law of wages implies.

The sphere of bargaining: In general. As the law of wages has been stated, it would appear that the rate of wages tends to settle automatically and without friction at a point which measures the marginal productivity of the workers. This is because conditions of perfect competition have thus far been assumed. But in practice the rate for any group of wage earners in any given industry is settled by bargain between employer and worker and fluctuates according to the relative bargaining power of these two parties to the wage contract. The actual rate in a chosen industry or for a chosen section of the laboring population may arrive at a lower point than that determined by economic law; on the other hand, it may rise above this point for short-run periods in isolated enter-

prises or localities. We cannot undertake to follow out in detail these complex variations, but it will be profitable to ascertain as nearly as can be the limits of this bargaining area within which actual wage rates fluctuate.

Wages above marginal product. Circumstances may arise in which the employer will voluntarily pay some of his workmen more than they are worth, thus leaving himself worse off than if he had refused to hire them at all. Employers are not devoid of feelings of compassion and do at times mix charity with business. Furthermore it may be to the positive advantage of an employer to accept a loss on the wage contract for a short time in order to avoid a greater loss involved in dismissing his employees. During periods of depression the enterpriser will look forward to a recovery of the market which will set his industry in full operation again. By keeping an experienced labor force together he will avoid the considerable loss involved in hiring new workers, breaking them in to the routine of the plant, and training them in his methods of management. It may be worth his while to do this even at the expense of paying his workers for doing nothing. These are but indications of the motives which may impel employers voluntarily to pay wages above the rate set by economic law.

It is also possible for the workers, under certain conditions, to force an employer against his will to pay them more than their marginal worth. By threat of strike a group of workers which holds a temporary monopoly on the supply of labor may compel the employer to make a choice between the alternatives of paying the members of this group more than the value of their marginal product or of suspending operations. If an employer could always close his plant without loss, it is not probable that he would ever accede to such demands. But this he cannot do. To suspend operations will at least wipe out his own wage income — the reward of his labor of management — along with the possibility of profits. If compelled to do so, he may pay wages above the marginal worth of the workers, surrendering to them a part of his own income up to the point where his share in the revenues of the enterprise has disappeared. Even here we do not have a final limit to the advance of wages, for it is conceivable that aggressive bargaining by the

workers may raise the rate so high that the enterpriser is not only not receiving payment for his own labor but is actually operating at a loss. Most employers have fixed costs which will run on unchanged even if they close down their industries. It is the lesser of two evils to run at a current loss, provided the loss is not as large as the total of the current fixed charges. This condition is of course self-limiting; its ultimate limit will be reached when the employer's equity in the business has been wiped out. But it is quite possible that an employer can be made to bear a loss of this kind, and there are plenty of cases of employers who have been forced into bankruptcy as the result of a series of losses on the wage contract.

We see then that there are three resistance points to a rise of wages. The first and most important is set by marginal productivity. It is rarely "good business" to pay a higher rate than this, and no employer will do so intentionally unless it be as a temporary expedient in abnormal circumstances. It is also rarely good business for laborers to demand more, since their employment is determined ultimately by profitable enterprise. The second check on rising wages is exerted at the point where the employer's profit and wages of management disappear. This is the maximum limit to the range of bargaining except in industries materially affected by fixed costs. The highest conceivable rate under any conditions is set at the point which causes a current operating loss equal to the loss involved in closing down the industry. Wages will rise above the first of these points, that determined by marginal productivity, only when the workers are possessed of some unusual bargaining advantage. Even then such a rise of wages can appear only in isolated cases and for short-run periods. When his wage cost rises to the second of our resistance points, the employer will soon give up the struggle. No employer will be content to work indefinitely without reward. The duration of a wage rate at a level between the second and third resistance points is automatically self-limiting by inevitable insolvency. We may safely conclude that advances of the wage rate above the point of the discounted marginal product will appear only as exceptional and short-lived occurrences.

Perhaps notice should be given to a type of situation sometimes

encountered in the business world which is often cited in refutation of this statement as to the temporary nature of wage rates above the marginal productivity level. Tightly formed and aggressive unions do sometimes raise their wage rates to extravagant heights and hold them there over considerable periods of time. But they do this by monopolizing and restricting the supply of labor. This raises its marginal productivity and drives up the price of the product on the market. The employer who pays these high rates does not necessarily pay more than the discounted value of the marginal product in the monopolistic situation which exists.

Wages below marginal product. We may now consider the reverse condition. As we have seen, modern industry requires such complex division of labor that there is no clearly defined test of the marginal value of the product of any group which all parties to the wage contract will accept as fair. Since this is true in individual cases, may we assume that these valuations result in some way from the operation of competitive forces in the general industrial world and that each employer adopting this rate will pay to the worker his maximum value to industry as a whole, even if this does not conform to his exact economic worth in the employer's own enterprise?

Such an assumption would rest upon the belief that competition works as freely on one side of the labor market as on the other. But this is untrue in actual life. For wage earners obviously are not as a rule in position to take full advantage of competition among employers. To do so they must be able to play off one employer against his competitors, to move about at will from one locality to another, to discard service in one industry and undertake employment in a different one whenever it is to their immediate economic interest to do so. All of this requires a degree of mobility in the ranks of labor which the average workman does not possess and which he is prevented from acquiring by (to mention but one factor among many) the mere cost of transportation for himself, his family, and his household effects. It must be remembered that the labor market ordinarily requires the workman to carry his labor to the employer. Perfect competition in the labor market then implies physical mobility among laborers, to attain

which they would have to overcome obstacles of financial expense, ties of home and community, custom, tradition, sentiment.

There seems to be no doubt that the bargaining weakness of the workers has often resulted in a wage rate temporarily or in particular situations lower than their marginal economic worth. Local situations may arise in which the play of ordinary competitive forces is interrupted. This condition of imperfect competition may enable employers to obtain labor for less than it is worth; *i.e.*, at wages less than marginal productivity. Such a state of affairs, in particular situations, may be long continued. Conditions of this kind have sometimes been disclosed when a compulsory increase in wages has not resulted in either unemployment or a higher price for the product.

It would aid us to limit the range through which the wage may fall if we could ascertain with precision lower limits to the sphere of bargaining comparable to the resistance points which check a rise of wages. In a general way two such resistance points can be marked out. The first of these is determined by the wage earners' traditional standard of living. Among the skilled labor groups especially, self-respecting workers will reject offers of employment at rates which spell a degradation of their standard of life and a loss of social status. A second and ultimate lower limit is reached when the rate falls to the minimum of subsistence. If starvation is to be their fate in any case, free men will prefer to starve at leisure rather than while working for another man's profit. Even if they were content to work for less than subsistence wages, the duration of such a wage rate would be limited by its inability to maintain the supply of labor.

Differences in wage rates within occupational groups. For purposes of exposition, we have thus far assumed that the workers in each specific line of occupation are so homogeneous that one can be substituted for another at will. Under this assumption all members of a given occupational group would receive the same rate of wages; all carpenters would be paid alike, all stenographers, all machinists, all school teachers, etc. Everyone knows that this is not strictly true in fact. The wage rate varies for different members within an occupational group and also for different

localities. Of course we are here considering, not merely the money payment, but the real wage. It is necessary to be certain that apparent differences in wages are real differences. If there are wide discrepancies in living costs between different localities, variations in money wages may not measure the real incomes of the workers. If employment is erratic, as in seasonal industries, a high daily rate may not mean large annual earnings.

Considering now the lack of uniformity in the real earnings of different members of the same occupational group, it is obvious that this fact may sometimes be due to relative difference in bargaining power such as we have discussed above. Women are habitually paid less for the same service than men. Negro labor in the United States is similarly underpaid in comparison with white labor, and the same is true of recent immigrants in comparison with workers who have acquired our culture and standard of living. The timidity of women workers, their inability to take advantage of potential competition among their employers, their indifference to opportunities to improve their immediate economic status weaken their bargaining power. In the case of colored labor, social prejudices and the improvidence of the negro, in the case of recent immigrants, their ignorance and contentment with a traditionally low standard of life go far to explain the fact that these groups do not receive the same pay for a given service as other workers. These factors influence wage rates indirectly through their effect upon the bargaining power of the workers.

But of chief importance in explaining lack of uniformity in the earnings of different members of the same labor group is the fact that no labor group is quite homogeneous in productivity. If we select any occupation at random, we will observe that all its members do not have the same productive efficiency. Noticeable differences appear in the number of tons of coal credited to the account of different miners at the end of an eight-hour day. The tenders of machines in a cotton mill will differ somewhat in the speed with which they perform their operations and in their ability to sustain a given rate of production through the hours of the working day. Other examples could be cited. Closer observation will show that these differences in productivity are greater in the

case of some labor groups than others. In the crudest forms of labor the difference in the output of one member of the working force as compared with another is usually slight. So also with the great army of workers in the factories; there is a marked uniformity in productivity among the millions of machine operatives in the great industries. The differences are great among skilled craftsmen, where native ability and training play a relatively large part in the mastering of the trade. And differences are greatest of all in the highest labor groups, the learned professions, and the superior grades of managers and superintendents.

Trade unionism and wages. Certain tendencies in the modern industrial world are working toward greater homogeneity of large labor groups and greater uniformity in the wage rate applying to these groups. Perhaps the most important of these tendencies is the growth of labor organization. Trade unionism flourishes especially among the remaining branches of the skilled trades, those groups of workmen which would disclose large individual differences in productivity if governed by the free play of self-interest. The essence of trade unionism is conscious and intentional coöperation for purposes of common benefit among the members of these groups. One standard device of the union is *collective bargaining* — the demand that employers hire the entire working force as a unit according to terms set in a single wage contract. Another standard device is the uniform rate, or the *union scale* of wages. Each employer who deals with the union is obliged to pay a standard rate to all members of the craft and to accord to each the same treatment in respect of hours of work. As the trade union grows in strength, this standardization of wages spreads from one industry to another until it covers the entire nation. Though not officially outlawed by the union, the practice of paying rates of wages higher than the union scale to selected individuals is frowned upon, and the public opinion of the group tends to prevent any member from accepting favors of this kind. Wherever the nature of the occupation lends itself to such practices, the unions tend to standardize the productivity of their members as well as their rate of wages. If their output is measured by units of product; *e.g.*, the laying of so many bricks per day, the con-

struction of so many moulds for iron castings, the turning of so many pieces on the lathe, organized labor will set up a certain number of these units as constituting a day's work. If the product is not subject to such enumeration, the same result is reached by setting a "decent" pace of work which becomes the norm for all good union men. The motives which lie behind these practices are complex, consisting chiefly of a desire to stifle individual competition which might be utilized by employers to disrupt the union, and to defend the workmen against the menace of unemployment. Without attempting to discuss the problem in all its bearings now, we may observe that the effect of such practices as these is to promote a tendency toward uniformity in productivity and earnings for large groups of workers over wide territorial areas.

Systems of wage payment: The time base. Where differences of individual productivity persist despite certain tendencies toward uniformity, the business world attempts to take account of them by adjusting the method of wage payment to serve the double purpose of increasing the reward of the more productive workmen and stimulating the individual's desire to increase his productivity. The perfecting of these devices of wage payment and labor management has led in recent years to the development of a new industrial science and a new profession for business executives. The traditional method of wage payment has been to base the rate of wages on a unit of time — the hour, the day, or some longer period. This method is difficult to adjust to individual differences in productivity and tends moreover to stifle the ambition of the better workmen. Employers who use the time base for wage payment generally attempt by disciplinary methods to induce from their workmen a rate of productivity at least as high as the wage rate. Slight differences in productivity are disregarded, and a "drive" policy of labor management is relied upon to keep the slackers up to the mark. In cases where differences in individual productivity are so marked that they cannot be overlooked, allowances are made for the superior efficiency of the better workers by means of differentials, which measure roughly their specific productivity as compared with the average man. These allowances do not always take the form of increases of money wages. The faithful and industrious

members of the labor force may be given privileges in the form of the more attractive jobs, and they will normally be the first to be employed and the last to be discharged in periods of business fluctuation.

Price rate wages: General description. The difficulty of adjusting time wages to individual productivity and the tendency of this system to impair the workers' efficiency have caused modern business executives to devise new systems of wage payment which will be free from these disadvantages. These systems vary in detail, but they all embody a common principle of *piece rate wages*. The first step in establishing a piece rate system is to resolve the product of the worker into a definite, measurable unit called the "job." The unit of measurement must of course be adjusted to the conditions of the particular enterprise. An iron foundry may adopt as a "job" a standard mould or casting; a coal mine, a ton of coal; cotton mills may use for this purpose a yard of their standardized product; the shipping department may have for one of its labor groups the construction of a box of standard size; for operators of automatic or power-driven machines the "job" may be a certain number of units of a processed product. The variations in detail are almost infinite, though the principle is the same in all cases. The next step, after settling on the "job," is to discover by experiment how many of these units the average workman can produce in a given length of time. This is usually done by selecting a subject for the experiment and timing him with a stop watch. With this information at hand, the final step is to fix a rate or price on the "job" which will return the average workman a rate of money wages equivalent to the customary earnings of men of equal skill in the general world of industry.

Under the piece rate system the workmen are invited to make the most of their opportunities, to produce as many units as they can in the course of a working day, and hence to prosper in accordance with their various capacities. Obviously this system of wage payment tends to break up the homogeneity of a labor group and to accentuate the difference in individual productivity and in the money earnings of workmen performing the same sort of labor. Refinements of piece rates, commonly known as *task and bonus*

systems, modify somewhat these tendencies to increase individual variations in productivity. These improved systems of wage payment also vary in detail, but many are alike in that they guarantee a standard rate of *hourly* earnings to all the workers, while rewarding those whose output exceeds the normal daily product by paying bonuses proportioned to the amount of the excess output.¹

Piece rates and the economic law of wages. These developments in the science of labor management do not upset the operation of the economic law of wages. In the first place, the piece rate system is applicable only over a relatively small range of industrial activity. It cannot be applied at all to workmen whose product is not tangible and definitely mensurable, and this is the case with by far the major part of hired laborers. Moreover even in those industries which have adopted the piece rate system the rate of wages is not divorced from the principle of marginal productivity, but more closely assimilated to that principle. The basic piece rate is frankly set to harmonize with the customary earnings of the group in question and is therefore governed by the general experience of the business world where marginal productivity is the norm of money wages. Though it results in differences within the shop which employs it, these differences are almost automatically adjusted to variations in individual productivity. The working force, instead of being merged into a common group with a single marginal productivity and a uniform wage rate, is broken up into smaller groups, each with its own peculiar productivity and its own rate of wages. The earnings of the general run of factory operatives will tend to conform to the general wage rate of their class, while the superior workmen will receive allowances which measure the excess of their products when compared with that of the average man.

Local variations in wage rates. Wage rates not only differ for different members of a given occupational group in the same place, but they vary between localities. To some extent these differences may be attributed to the effect of the bargaining factor which we have discussed. For various reasons the bargain which fixes the

¹ For more detailed exposition of the piece rate and task and bonus systems of wage payment see C. B. Going, *Principles of Industrial Engineering*, 1911.

actual wage rate within the limits described on a preceding page terminates variously in different localities. Enterprisers are not alike in their efficiency as managers, and consequently different business concerns vary in profitableness. Some units of a given branch of industry are equipped with better machinery or possessed of more favorable markets or otherwise favored as compared with competing units. These differences affect the ability of an enterprise to pay wages, and they also affect the aggressiveness with which the employer bargains for a saving of cost in the wage contract. Employers with the smallest profit and those operating without profit or at a loss are proverbially the most aggressive bargainers. On the other side of the labor market, differences in the character of the working people, in their standards of life, in the degree to which they are unionized, and in other local conditions materially affect the bargaining strength of different groups of laborers. Such differences in local conditions modify the resistance points in the fluctuation of wages and cause local variations from the norm. The effect of the bargaining factor upon local wage scales is much more noticeable in the case of some types of labor than in that of other types. The wages of women in any given locality are always much affected by their lack of bargaining aggressiveness; hence these local wage rates are apt to vary markedly as between different business centres. Illustrations of these local variations of the wage rate for a given type of labor can be found in any statistical summary of wages paid in different cities throughout the country.

The basic cause of local variations in wage rates however is the fact that the labor market is broken up into small fragments between which the forces of competition either do not operate at all or operate haltingly and with much friction. While prices of commodities in general in different places cannot differ by more than the cost of transportation between these places, labor is an exception to this rule on account of its extreme immobility. If all the workers in a given occupation could and would move at will over the whole industrial field, local wage rates would tend toward much greater uniformity. But in reality each business centre forms a labor market of its own, where the relation of supply to demand

may be quite different from that of other centres and where consequently the marginal productivity of the workers will also be different.

Local wage rates tend to vary in rough proportion to the cost of living; they are lower in small towns than in the metropolis, in Southern than in Northern cities, in the Middle West than on the seaboard. This fact lends support to an erroneous theory that the wage rate is determined by living costs. But local differences in the cost of living can affect wages only through their influence on the distribution of the labor supply and on the bargaining power of the workers. The cheap living which they enjoy in a country town or in a small manufacturing city not in close communication with the general business world may be an element in the unwillingness of a labor group to move to the metropolis where wage rates are higher. In this case the low wage rate will measure a low marginal productivity. Low living costs may also perpetuate a traditional wage scale in a given locality which is possibly lower than the economic worth of the workers in that place. Contentment of the workers with a customary mode of life weakens their aggressiveness as bargainers and, in conjunction with their unwillingness or inability to move elsewhere, operates to allow the local wage rate to remain low.

In concluding this topic it should be observed that modern developments are tending to reduce these local variations in wage scales by broadening the labor market and removing differences in the bargaining power of different labor groups. Cheap transportation increases the physical mobility of the workers; newspapers and magazines with nationwide circulation stimulate similar desires in widely separated labor groups and stir the ambition of the more backward; education breaks down the barriers of provincialism and traditional standards of work and pay. Negro labor is beginning to migrate to the north; the small towns and back country cities witness an annual exodus of their younger laborers to the chief business centres; immigrants quickly assimilate our standard of living, and their children appear to be born with the American's voracious demand on life and his restlessness to better his condition. The spread of unionism promotes the

mobility of local groups and at the same time stiffens their bargaining power. The peculiar weakness of female workers is disappearing as it becomes an established custom for women to seek a life of economic independence and a permanent place in the industrial world. These forces and many others of similar nature broaden the area within which a common wage rate holds sway.

Discounted marginal productivity is fundamental. Because of the complexity of the labor market it has been necessary to introduce various qualifications into our law of wages in order to bring it into rough conformity with the facts of industry. We have recognized the fact that the actual wage rate is determined by bargaining and have marked out the limits of the sphere within which the incidence of the bargain will fall. We have taken account of the existing lack of uniformity in the wage rate for members of a given occupational group and have accounted for this by referring to differences in bargaining power and especially to differences in individual productivity. We have also noted local variations in wage rates for a given type of labor and have shown that these result from local differences in relative bargaining power and especially from maladjustment in the labor supply.

But this recognition of complicating factors must not be permitted to obscure the essential truth embodied in the economic law of wages. The wage bargain revolves around the point of discounted marginal productivity as a norm; actual wage rates cannot long persist above this point, and all the developmental tendencies in our society work against their falling far or remaining for long below this point. Homogeneity of labor groups is promoted by various tendencies which we have noted; other forces are broadening the scope of the labor market and promoting a more even adjustment of the local supply of labor to the demand for it. We may conclude that no device of wage payment and no alterations in the method and procedure of the labor market can materially affect the fact that the laboring class in a society sanctioning private property and individual enterprise will normally receive wages approximating the value of their specific contributions to society's income. If they are members of groups whose individuals

have much the same rate of productivity, the value of each man's contribution will be equated, through the principle of substitution, to that of the marginal man in the group. If they must receive their share of the social income before the sale of their product, their money incomes will be reduced by the discount process through which present and future values are equated. Upon these valid principles rests the economic law of general money wages, the law of the discounted marginal product of labor.

Other types of labor income. The foregoing law of wages was developed with reference to the hired worker; but its conclusions will serve to explain the money incomes of those who are their own employers. Here the typical case is complicated by the fact that the independent worker must almost of necessity be an owner of capital and to some extent a manager of enterprise. The physician, the lawyer, the barber — examples could be multiplied indefinitely — possess productive capital of one sort or another and often land as well; their money incomes are a mixture of rent, interest, and wages, no single element in the income being clearly defined in contrast with the rest. It is not impossible however to arrive at a rough measurement of that part of their income which is attributable to labor, by deducting allowances for capital and land at the market rates of interest and rent.

In the case of independent workers, no intermediary stands between them and the market. Their money incomes are drawn directly from the buyers of their products. Their incomes rise with a rise in the price of their product and fall as prices fall, tending to settle at any given time at points which measure the marginal utilities of their services to consumers. The fact that these groups are less homogeneous than are groups of hired workers, that there are for example relatively large differences in the productive efficiency of physicians, attorneys, and consulting engineers, tends to bring about marked variations in earnings within these groups. But apart from these variations in money wages, these classes of the laboring population constitute no real exception to the law of wages. The normal income of the doctor will be high or low in harmony with the relative scarcity or plenty of the services of this profession or, to use the terms of our law of wages, in harmony

with the marginal productivity of the group. So it is with other classes of independent workers.

The more apparent exceptions to the law are exhibited by those groups of workers whose incomes are determined in some material degree by non-economic considerations. Such are those connected with charitable enterprises, members of religious orders, to a less degree teachers and government officials. In many cases the services of these groups are not sold to consumers at prices which measure their marginal utilities to the buyers. There can be no direct relation between the earnings of such workers and the market prices of their products, for the sufficient reason that the prices are not determined by market considerations. What value the consumers would place on the services of men engaged in gathering and distributing the mail we can never know until this business becomes competitive and makes its bid for patronage in rivalry with other outlets for the purchasing power of the consuming public. So it is, though to a less degree, with the school system. We can only guess how many teachers would be employed and what money incomes would be returned to them if these things were left to the test of the marginal utility of instruction to the people as a whole.

But even here the economic law of wages is by no means without effect. These occupations must compete for workers with opportunities in the general world of business and professions. The rewards offered cannot be set too far below those afforded by competing occupations to men and women of the same ability, else the former enterprises will be unable to obtain the requisite supplies of labor. The periodic complaint over the dearth of teachers, ministers, soldiers in the standing army, and "good" men in all branches of the civil service shows what happens when the wage policies of these non-commercial enterprises are out of line with the general economic conditions of the times. Making allowance for the non-material rewards of such types of labor as we have been considering, such as prestige, security of tenure, relief from strain of competition, congenial surroundings and associates, and the like, we shall find that wages in these occupations show rough similarities to the ruling rates fixed by economic forces in comparable lines in the business world.

EXERCISES

1. The annual income realized by an automobile repair shop with different numbers of mechanics employed would be as follows :

<i>Mechanics</i>	<i>Income</i>
4	\$45,200
5	51,000
6	55,200
7	58,100
8	60,000
9	61,200

How many mechanics would be employed at a wage rate of \$150 per month?

2. Men's shoes are sold in a retail store at a gross profit or "mark up" of \$1.00 per pair. The salaries of salesmen are \$160 per month. Assume that the following relationship exists between number of salesmen employed and monthly volume of sales :

<i>Number of salesmen</i>	<i>Volume of sales</i>
1	480 (pairs)
2	880
3	1,200
4	1,440
5	1,600
6	1,680

- (a) Determine the marginal cost per unit of sales when different numbers of salesmen are employed.
- (b) How many salesmen would be employed and what would be the volume of sales?

XIX

DIFFERENCES IN WAGES. WAGE PROBLEMS

Differences in wages: Social stratification of the laborers. The tendency toward uniformity of wages in any occupational group, as developed in the preceding chapter, rests upon the fact that within certain limits labor is fairly homogeneous; *i.e.*, that the workers are in actual or potential competition with each other. We have taken account of certain obstacles to competition within any given occupational group which cause differences in the wages paid for that type of labor. But the great contrasts in wage income which one observes in the world of affairs have to do with the incomes of people in quite different occupations, for example those of the business executive and the manual laborer. Such great differences of income are evidence that the laboring population as a whole is divided into social strata whose lines of demarcation are sufficiently rigid to remove their respective groups from competition with each other.

Many attempts have been made to classify the gainfully employed people of the nation along these broad lines of social stratification. One such classification, sufficiently exact for our purposes, divides all labor into five fairly self-contained groups.¹ At the bottom of the scale are the unskilled workers in all occupations — the manual laborers of the farms, the railroads, the industries of the country. In the second group are the manual workers of some little skill whose occupations require a certain amount of responsibility and mental alertness — the machine operatives of the factories, the pick miners, the motormen of the street railways, and the like. The third group comprises the skilled craftsmen, the aristocracy of the manual workers, such as the craftsmen of the building trades, the skilled workers of the steam railroads, the die and pattern makers of the metal trades. In the fourth group are found the “white collar” workers who provide the office and clerical staffs of

¹ F. W. Taussig, *Principles of Economics*, 1916, Vol. II, pp. 134–137.

business enterprise, the sales forces of retail establishments, teachers of the lowest grades. The fifth and highest group contains those superior workers who are either self-employed or occupy salaried positions of heavy responsibility. The learned professions, the managers of enterprises, the higher public officials, the independent enterprisers are examples.

Wages of each group tend toward uniformity. To the extent that hard lines of social stratification separate these groups, preventing a transfer of membership from one to another, their normal wage rates will diverge. Within each class however there is a tendency for the earnings of the different sub-groups or specific callings to become fairly uniform. This is particularly true of the wage rates within the first two classes. The unskilled laborer may today be a digger of ditches, tomorrow a harvester of crops, later a street sweeper, hod carrier, roustabout, janitor. The rank and file of machine operatives are similarly transferable as between different jobs.

The other great social groups differ in this regard from the two lowest strata chiefly in the length of time required to iron out discrepancies which appear. The third group in our classification falls more definitely into separate skilled trades or callings. The workers in this group at any given time cannot shift from one occupation to another with the same ease as the members of the lower groups, for each specific calling requires a fairly high degree of technical training, which, once acquired, tends to commit the individual more irrevocably to his chosen occupation. Bricklayers must be content to remain bricklayers, machinists to remain machinists. Hence there is no immediate transfer of the existing workers such as otherwise would occur if the relatively high earnings of one of these callings made it temporarily more attractive than the others. But the distribution of the labor supply which tends to remove discrepancies of this sort is only delayed. The skilled crafts call for very little preliminary education of a general nature, and hence the yearly supply of newcomers qualified to learn the trades, youths who annually assume the duty of self-support, is relatively large. These new workers tend to equalize the numbers in the different trades in proportion to the demand for their services, by applying in in-

creased volume for admission to the trade which for the time being is unusually attractive.

Similar forces may be observed at work within the two upper classes of the working population. The great mass of the workers in our fourth classification; *i.e.*, the subordinate "white collar" employees, clerks, stenographers, bookkeepers, and the like, though it falls into a large number of slightly dissimilar occupations, is rewarded with money incomes which vary within fairly narrow limits. The workers in this group can shift from one specific calling to another with greater ease than can the skilled craftsmen, although a longer preliminary period of education is required to equip the youth of the country for this type of service in industry. Commercial schools, "business colleges," night schools, correspondence and extension courses bring the rudiments of the specialized knowledge of business routine within the reach of great sections of the middle and lower middle classes. One of the impressive results of the modern educational system is the tremendous increase in the proportions of young men and women who elect to seek their livelihood in the "white collar" group of workers. These new increments of labor seek out the lines of occupation which offer the highest rewards, thus tending toward an equalization of the membership of the various sub-groups of this class in proportion to their respective demands.

Least uniformity in the highest group. In the highest class uniformity of earnings is least marked. Here not only is the individual worker committed permanently to his chosen profession by reason of the long and expensive preliminary training required to master a new one, but the annual influx of new workers is small as compared with the lower social groups. Family tradition and other matters of sentiment play a relatively important part in determining the distribution of labor among the professions. Innate differences of native ability and aptitude also exert their influence. Adjustments of numbers to differences in earnings are longer delayed and less automatic. Yet one who observes the succeeding generations of college men at the business of choosing their careers cannot be unaware that the recruiting of new workers, even in the learned professions and in lines of activity which are comparable in the

degree of preliminary training they require, is much influenced by the matter of relative earnings.

The independent enterprisers of the business world, who belong to this general social group, are to some extent in a class apart. Here success depends more upon inborn qualifications and less upon training. At their higher points enterpriser's incomes are subject but little to the levelling tendency of competition among workers which conduces to uniformity of earnings between other lines of occupation. This is much less true of the lower levels. The salaried employee of the highest grade, the young lawyer in the employ of a firm of attorneys, the sales manager, production manager, consulting engineer, industrial chemist, has ever before him the opportunity to venture forth as an independent enterpriser in his chosen calling. Conversely the independent business man may give up the struggle for profits and drop into the more secure position of salaried employee almost at will. These transfers from one position to the other occur when business profits in their lower ranges and the scales of salaries for workers of commensurate ability diverge in any marked degree.

Wages and the contrast of social classes. The great differences in wages which we observe in actual life mark the contrast in economic position of the several social classes. A large railway system usually gives within itself a picture of the whole range of social classes from top to bottom. There are the president with a salary of fifty thousand a year or more, the managers of various departments with perhaps half this money income, the skilled workers on the trains and in the repair shops with average earnings of about twenty-five hundred dollars a year, the office workers with much the same rate of income, the maintenance of way men receiving little more than a thousand a year, or one-fiftieth of the president's salary.

It is often remarked that the greater differences in wages merely indicate difference in the social importance of the functions performed by the various social classes. As a type of social service, the task of driving a giant engine on a modern railway system appears to be more important than the pick and shovel work of the section gang. The work of the architect who plans and supervises

the erection of a beautiful building seems more useful to society than the driving of nails by the carpenter or the lifting of burdens by the hod carrier. That the function of the surgeon is of greater social significance than that of the hospital's janitor appears to be clear beyond all question. This "common sense" view of the matter is often advanced as proof that, despite great inequalities, men after all are rewarded in proportion to their deserts.

But we must be on our guard against drawing from such comparisons of individual workers inferences as to the relative importance of functions. Each function is indispensable to production. Imagine a railway system manned entirely by presidents and superintendents, without the labor of engineers, firemen, and trainmen; or a hospital staff composed solely of surgeons with no nurses, orderlies, cooks, or caretakers. When one seeks to evaluate the different productive functions in their entirety he is on dangerous ground, just as when one attempts to determine the relative importance of land, labor, and capital as separate factors of production.

As a matter of fact, whatever may be the relative importance of the several social classes, it is not this which determines the rate of wages received by the members of each, any more than relative social importance determines the prices of bread and diamonds. Here exactly as in the case of prices in general, we are dealing with marginal utility. In the world as it is, the basic cause of these great contrasts in wage incomes is the distribution of the working population among the various social levels. The economic forces in the labor market do not set values on functions but on individuals; it is the indispensability of the individual which determines his share in the social income. This indispensability is determined in turn by the number of available substitutes for the individual; it is the marginal worth of labor groups that counts. The railway president commands a high price because the number of men capable of performing his function is so very small. The hod carrier's wage is small because he is only one of a vast number of available units.

Permanence of cleavage between social classes: The facts. That great differences in wages persist through succeeding generations suggests that some obstacle prevents the free flow of labor from one social class to another. If the youth of each generation were entirely

free to choose their occupations there would be an immense increase in the number of radio singers and movie heroes and a great dearth of farm laborers. Wages at their higher levels would fall, and that class which now occupies the lowest station would rise to the top of the social scale.

Wages in different occupational lines would not be entirely equalized, even if there were no barriers between the social classes. The distribution of labor among these different occupations would not be precisely in proportion to the demand for their respective services. Some jobs would prove more attractive to the average man than others; perhaps because of the more cleanly work involved, the pleasanter surroundings, the greater independence, perhaps because of differences in the length of time taken to master the details of the job. The exhausting task, the dirty and unpleasant job, the one which involved risk to life or health would be shunned by the workers, and a flood of labor would flow into the more attractive lines of occupation. Coal miners, being relatively scarce, might well get more than managers, and ditch diggers, for the same reason, more than bank presidents. But these differences in wages would merely serve to compensate for the greater irksomeness of some tasks as compared with others. In other words, given perfect mobility of labor, wages would be equated to the *real costs* of the different jobs, the human exertion and sacrifice which they entailed. The law controlling these differences in wages would be the same law as now operates; namely, the law of marginal indispensability. But today some barrier to the flow of labor from one social class to another breaks this correspondence between differences in wages and differences in labor cost; indeed as things go in the modern world, it is most frequently true that the man who works in dismal surroundings at the most disagreeable toil is the poorest paid of all.

Explanation: Nature and nurture theories. Speaking broadly, two explanations are offered for the persistence of social stratification through the ages. There is the belief that conditions of birth and the heritage from his ancestors determine the rank which a man will take among his fellow men. Those in the humblest station are predestined at birth to occupy this position in the lowest income

group. Those who rise to higher levels do so by virtue of inborn characteristics which can neither be produced intentionally nor balked of their reward when present. According to this view of the matter heredity is the controlling factor in the destiny of men; the present stratification of society is a natural order determined by nature's laws.

In contrast with this view stand the environmentalists, who assert that man's social destiny is determined after birth by the forces which play upon his plastic consciousness during his early years. In their opinion the children of wage earners fail to rise above their fathers' station in life because of the stunting of mind and body and the impoverishment of will which they suffer from the unwholesome environment of their childhood. Able men are born in about equal proportions in all ranks of society, but those in the lowest ranks are the victims of physical, moral, and intellectual malnutrition. They are denied opportunities for self-development, opportunities which are showered in abundance upon the children of the rich, regardless of intrinsic merit.

For the student of society the chief significance of this search for the basic cause of social stratification is its bearing on the possibility of reform. It may well be that the stability of the social order, to say nothing of the welfare and happiness of the average human being, would be improved if life conditions within the social system were more nearly uniform. One's attitude toward the possibility of reducing these contrasts through conscious social effort will rest upon his acceptance of one or the other of the theories of social stratification discussed in the preceding paragraphs.

It is not pertinent to our study to follow out in detail the issue joined by the nature and nurture theories, but we should note that the question is not to be set at rest by the superficial evidence usually advanced in favor of one position or the other. The environmentalists do not deny that men differ greatly in inherent ability, nor that these differences are present at birth. The question however is this: Does nature distribute these favors of natural capacity with regard to the artificial lines of social stratification — lines determined primarily by the distribution of wealth? Among a given hundred thousand children, one rare genius may appear, a

very few will be born with great innate talent, a somewhat larger number with ability beyond the ordinary, the vast majority will be mediocre, and a small residuum will be composed of mental, moral, and physical weaklings. Do these proportions hold good for all income groups in society, or is the proportion of genius and talent higher among children born of well-to-do parents than among those in the lower labor classes? This will always be a moot question until it is possible to separate the two factors, heredity and environment, and observe the operation of each alone. Our institution of inheritance is alone enough to prevent any such experiment. The children of the rich may or may not inherit the abilities of their father, but we know that they do inherit their fathers' wealth. So long as all children of different parentage are not treated equally nor equally provided with opportunity for self-development, it will be impossible to demonstrate conclusively which of these two factors plays the major rôle in the stratification of society.

Whatever the ultimate cause of the stratification of society into rigid non-competing social groups, there can be no doubt that the stratification exists and that the different social classes are recruited chiefly from within. Their numbers do not expand and shrink in response to increases or decreases in the rate of wages which represents the average earnings of their various sub-groups. Striking cases can be cited of the rise of individuals to positions of prominence and wealth from the humblest beginnings; but the class from which these men sprang remains, as a class, what it is and has been — the most numerous and the poorest paid social group. As long as this is true, differences in wages will bear no necessary relation to the real costs of the workers.

Real wages. In any analysis of wages intended to show the comparative well being of the laborers at different times or places, it is essential to take account of differences in the level of prices. Our discussion of wages thus far has been devoted chiefly to money wages. But it is evident that high money wages may be neutralized by high prices and *vice versa*. To measure the well being of the laboring class, wages must be expressed in terms of the quantity of goods which the money will buy. When thus expressed, the term *real wages* is used, as contrasted with money wages. The adjust-

ment of money wages at different times or in different places to a common price level is easily made by means of an index number of prices, either of prices in general or of the prices of the most important commodities and services entering into the typical working-man's budget.

Changes in wages: The law of proportionality. For society as a whole, improvement in real income can be due only to some general development affecting the physical productiveness of society's economic system. But it does not necessarily follow that an increase in the total income of society will be evenly spread over the several shares in distribution. It is quite possible that the forces which bring about increased general prosperity may so operate as either to increase or decrease the fraction of the whole which goes to labor, as well as the average real wage rate.

For example let us suppose that the number of the laboring class doubles, while the land area, the supply of capital, and the arts of industry remain unchanged. The total product of the community will certainly be larger than it was before, for each additional member of the working population will have brought a pair of hands with which to aid the productive efforts of the group. But the marginal product of the laborers will be smaller now than it was before their number increased, since the productive factors which coöperate with human labor have remained fixed in quantity. There will be less land and less capital for each man to work with. Those laborers who apply their labor directly to the exploitation of natural resources must now work on less fertile lands, less productive mines and forests, or else upon an intensive margin of cultivation which has been forced beyond its former level. In either case the principle of diminishing returns will have reduced the product attributable to the individual laborer. In the factories, on the railroads, in other branches of industrial activity, a similar occurrence will have taken place. The additional members of the industrial population will have been compelled to divide with their fellows the use of the fixed supply of capital instruments; each will receive less effective coöperation from these instruments, and hence their marginal productivity will have fallen.

The aggregate income of the working population will not through

this cause become smaller than it was before. It is probable that it will be increased or, to state it otherwise, that the total share in the social income going to labor will be larger in size than it was before. But it is obvious that the average real income of the workers will have declined as a result of the forces which have caused the total social product to expand. Under the conditions assumed, it is apparent that the owners of land and capital will not have shared the fate which has befallen the laboring population. When the intensive margin of cultivation is advanced and the extensive margin broadened to include lands and other natural agents of inferior productiveness, we know that the rent of land rises. When each unit of capital instruments is used more intensively, its productive effectiveness is increased. The marginal productivity of capital throughout the whole field of industry will have risen, and the suppliers of loan funds will reap the benefit of this change in the form of a higher rate of interest.

We reach the conclusion that the change which we have assumed would have the effect of increasing the average real income of landlords and capitalists and of decreasing that of the workingmen. The same line of argument would demonstrate that an increase of land and capital, while the laboring population remained stationary, would tend (1) to increase the total social income and (2) to decrease the average real income of landlords and capitalists, while increasing that of the workingmen.

The foregoing hypothetical case brings to light a principle of universal truth with regard to the relative economic prosperity of the different income receivers in society, which principle may be formulated as follows: *any social change which reduces the ratio which one factor or combination of factors bears to those coöperating with it will tend to raise the average income of that factor or combination of factors as compared with the others.* If population increases faster than land and capital, real wages tend to fall in proportion to interest and rent. If capital accumulates faster than the increase in population and natural resources, wages and rent tend to rise in proportion to interest. If the available supply of land and natural agents expands out of proportion to labor and capital, the rent income tends to shrink as compared with wages and interest. This

generalization is called the *law of proportionality*. It is applicable not only to changes in a factor of production taken as a whole in relation to the other factors, but also to changes of balance among the sub-groups of a given factor. If within the labor supply, for example, managerial labor increases in quantity while other types of labor do not keep pace with it, the average real income of the one will tend to fall proportionately to that of the others. If one type of capital can be kept from increasing while all other types are expanding, the relative income of the constant type will increase.

Labor and the law of proportionality. This matter of proportion among the subdivisions of a productive factor is of particular importance in the case of labor. The fact is that any type of labor can increase its relative real wages by limiting its numbers or, having established a limit to its numbers, by limiting also its productive effort. If each particular labor group consumed its own product, its material welfare would be increased by any force or combination of forces which multiplied the physical productivity of its members. But this we know is not the case in modern society; the standard of living of one of our occupational groups is made up to but slight extent of the products of its own toil. The workers in the shoe factories are much more concerned about the supplies of foodstuffs, clothing, dwellings, coal, and other things, than they are about the supply of shoes. Each group in short is more concerned with the productivity of other groups than it is with its own.

This feature of our exchange economy has far-reaching effects which are not always favorable to general social welfare. It means that it is often to the immediate economic advantage of each separate labor group to restrict its membership and raise the value of its marginal product. Many devices may be adopted to produce this result. When the workers in a given occupation are organized into trade unions and are able to prevent those who are not union members from engaging in the occupation, they are tempted to adopt restrictive practices which artificially limit their numbers and, in an equally artificial manner, increase the numbers in other lines of activity. This will raise their marginal productivity, make their product relatively scarce as compared with other products, and increase their real wages. Much the same result is obtained if

the group in question adopts "make work" policies which set a limit to the output of the average man in a given time. This may be done by deliberately lowering the rate of output below the normal capacity of the workers, opposing labor-saving machinery, destroying the effectiveness of waste-reducing devices, and in many other ways. To be effective this decree must of course be accompanied by limitation of the numbers of the labor group.

It is not difficult to prove that society as a whole is always injured by any device which artificially reduces the productivity of any of its classes of workers. But no single group of workers is interested in the economic welfare of society as a whole; it is anxious about its own rate of real wages. Now the fact is that each labor group is injured by such restrictive policies as we are discussing when they are practiced by others but is benefited by the same policies when practiced by its own members. To be sure there are limits beyond which this practice cannot be pressed by the group without causing it to react indirectly against its members. If the services of the group become too scarce and of too high value, substitutes will be found which may displace them entirely. And of course if all groups of workers begin to employ such devices at the same time, they will checkmate each other, and the outcome will be general impoverishment for everybody concerned. Nevertheless restrictive devices not only result naturally from the existing arrangements of the social order but are often sensible and "practical" from the standpoint of the group in question, however injurious they may be to the laboring class as a whole and to the general social welfare.

Historical changes in the rate of wages. Differences in the general welfare of the industrial populations of different countries, or of the same country at different times, are to be explained in the light of the basic determinant of wages; that is, the marginal productivity of labor. There have been epoch-making changes of this character in the recent history of the world. The period since the beginning of the Industrial Revolution has been one of revolutionizing discovery of new sources of power, rapid accumulation of capital, progress of invention, improvement in transportation, perfection of the science of business organization and management. In 1801 the total population of England and Ireland was approxi-

mately 16,000,000; by 1901 it had risen to 41,500,000. At the earlier date the economic condition of the mass of people in England was one of appalling poverty and misery; during the century, despite the rapid multiplication of people, real wages rose steadily. The following table traces this rise of average real wages from 1830 to 1900, using the average of 1900 as a base.¹

<i>Year</i>	<i>Index of real wages</i>
1830	45
1840	50
1850	50
1860	55
1870	60
1880	70
1890	84
1900	100

The average real income of labor in the United States has always been higher and is now higher than in older countries less abundantly supplied with natural resources. The steadily rising wage income of the first three quarters of the nineteenth century must be attributed principally to our expanding frontier, which opened up seemingly inexhaustible supplies of land. This increase in coöperating factors permitted an unparalleled expansion of numbers — from 3,929,000 in 1790 to 38,558,000 in 1870 — while allowing the marginal productivity of labor steadily to rise.

A change in basic factors occurred during the last three decades of the century. During this period the frontier was gradually extinguished, but there was a rapid increase of capital, a continuous progress of invention, an extension and perfection of the transportation system, and a revolution in industrial organization and the science of management. The increasing factors were primarily capital and entrepreneurial ability. Expansion of these factors in other countries, together with some discovery and exploitation of formerly isolated natural resources, contributed through the extension of foreign commerce to improve the economic status of the American laboring population. During these thirty years, the number of people increased from 38,558,000 in 1870 to 75,995,000 in

¹ From A. L. Bowley, *Statistical Studies*, 1904.

1900. The rise of average real wages is shown in the following table in which the average wage of 1890 is used as a base.¹

<i>Year</i>	<i>Index of real wages</i>
1870	68.7
1875	72.5
1880	82.8
1885	98.2
1890	100.0
1895	102.0
1900	104.5

The period since 1900 has shown different trends of the wage income in different countries. For the world as a whole population continued to increase, but free supplies of land and natural resources disappeared, and the maintenance of a rising wage income depended upon increases and improvements of capital and business management. In many industrial countries there is reason to believe that the wage income began to fall between 1900 and the outbreak of the World War. The following table, which uses 1914 as a base, shows the trend of real wages in the United States from 1899 to 1933.²

<i>Year</i>	<i>Index of real weekly wages</i>
1899	99
1904	99
1909	103
1914	100
1919	101
1921	109
1925	124
1927	126
1929	130
1933	107

Especially noteworthy are the fairly stable levels of real wages during the period prior to 1920, the extremely rapid rise between that year and 1929, and the equally abrupt fall of real wages during the subsequent business depression.

The wide fluctuations in real wages in the United States during the disturbed times which have followed the World War are shown

¹ From T. S. Adams and H. L. Summer, *Labor Problems*, 1905, p. 514.

² From C. R. Daugherty, *Labor Problems in American Industry*, 1933, p. 185.

in the following table. The base year is 1929, and the figures refer to labor employed in the four largest industries : mining, manufacturing, construction, and transportation.¹

<i>Year</i>	<i>Index of annual real earnings</i>
1918	85
1921	67
1929	100
1932	44
1935	67

Wages and the standard of living: The workers' theory. If it can be said that the workers themselves have a theory of wages, it is that wages either are or ought to be determined by the laborer's standard of living. It is not to be expected of course that any theory or explanation which arises from the thinking of the mass of men in industry should have been worked out with scientific accuracy. For the most part, the wage theory of the workers lays stress upon what is fair and just rather than upon what is economically possible; in other words, it falls in the domain of ethics rather than in that of economics. But the standard of living theory is often advanced as an attempt to explain things as they are; as such it vies with the economist's theory of marginal productivity and must be considered in any study of wages.

In its most extreme forms the standard of living theory encourages extravagant expenditure on the part of the working classes. Wage earners are prone to oppose the efforts of social workers to reduce their cost of living by means of frugal purchasing of supplies, cheaper diet, thrifty budgeting of costs and income, and similar devices. It has been their belief that any gains resulting from these expedients would be appropriated by the employers in the form of a lowered wage rate. In the reverse direction many labor leaders and individual members of the rank and file have favored the inclusion in the workers' normal standard of living of expensive items of consumption. The idea here is that such extravagances, as soon as they become habitual, will play their part in determining the rate of wages. Underlying this whole theory, we usually find unexpressed the belief that the employer's remuneration of his hired employee is

¹ *Ibid.*, Revised Edition, 1938, p. 153.

determined either autocratically or at any rate arbitrarily and that it tends to gravitate toward some cost-of-subsistence minimum, where its decline is arrested either by the force of sentiment (public sentiment or the employer's compassion) or by the sheer impossibility of obtaining a supply of workers for a smaller wage.

The true effect of the standard of living. The fallacy in the crudest of these corollaries of the standard of living theory is obvious. But there is a real relationship between the standard of living and the wage rate, which should be understood by students of economics, especially in these times when it is popular to attempt to fix wage rates with reference to the cost of living. There are two ways in which a high standard of living may be effective in raising the wage rate or preventing its decline. In the first place, a firmly entrenched standard of living may operate to increase the bargaining power of the workers and enable them to resist a reduction of wages or to win an advance. To have this effect the worker must be so firmly attached to his standard that he will refuse to work for a wage which involves a degradation of this traditional mode of life. When this is true, the standard of living sets the lower limit to the range of wage bargains; a gradually rising standard will elevate this lower limit and cause the actual rate to approach more closely and adhere more permanently to the point determined by marginal productivity. In view of what has already been observed on this subject, it need scarcely be pointed out that the wage cannot be maintained by these means at a point above the marginal worth of the workers for any considerable length of time. Nevertheless there are on record enough cases of practical success in winning increases in wages to lend empirical support to the argument that the actual wage rate may be influenced by the standard of living.

In the second place, the standard of living must be viewed as one of the forces determining the marginal productivity of the workers, in which character it plays a part in the complex of economic forces which fix the normal rate of wages. The reason for this is two-fold: the standard of living affects the productivity of the individual workmen, and in a fundamental way it also exerts control over their numbers. It is common knowledge that a close correlation exists between the productive efficiency of the worker and those phases

of his standard which consist of physical maintenance. More nourishing food, better clothing, better housing conditions improve the physical and mental alertness of the workingman and resist the destructive effects of fatigue. We are also beginning to see that less tangible elements in his mode of life have a similar effect upon the worker. Leisure for intellectual and spiritual development, freedom from worry and anxiety, a hopeful prospect of social advancement for his children, all contribute to the worker's effectiveness as a producer. The argument in support of the "economy of high wages" rests in part upon such considerations as these, claiming that, when once the wage rate of the low-paid worker has been raised, he automatically becomes worth the higher wage. We need not push our analysis of this phase of the standard of living theory very far in order to appreciate its truth and also to grasp its limitations.

But the wage rate of an individual member of a labor group is governed, not solely by his own physical fitness and his own industry, but also by the bare fact of numbers within his group. The standard of living operates over long-run periods to set a limit upon the numbers of workers in the different occupational groups, as well as to the total number who seek employment in industry. Viewed from this broad social standpoint, the standard of living may be defined as that quantum of physical and psychical enjoyments which the individual demands in preference to marriage and the procreation of children after marriage. In this day of deliberate postponement of marriage and equally intentional control of the birth rate, the importance of this factor as a check upon the growth of population will be granted without much argument. By advancing its standard of life, a given labor group not greatly subject to infiltration from other classes of workers may make itself progressively a scarce factor in industry, raising its marginal productivity over a long-run period of time and advancing its rate of real wages.

For certain wage-earning groups this implication of the standard of living theory has come to have very practical meaning. The higher grades of workers began long ago to give evidence of the influence of their standard upon their rate of increase. As a result

it has become a commonplace among social students to attribute to this influence some of the credit for that rise of families in the social scale which has characterized society since the Industrial Revolution. Unfortunately the more necessitous groups, whose poverty is due to their excessive numbers and a consequently low marginal productivity, are least affected by this factor. It is among the classes which are already prosperous that we find the greatest inclination to elevate the standard of living in preference to an increase in the birth rate. As a result the principal influence of the standard of living, viewed as a check upon numbers, has been in the direction of maintaining existing high wages and over long-run periods slowly advancing them, rather than of raising the low wage rates.

Wage regulation: By action of employers. There have been many proposals in recent times to fix the wage rates of employees according to dictates of social expediency or moral principle. These proposals are usually intended to raise the income of the wage earners, particularly the manual workers, above the levels set by competition. In some cases individual employers have adopted wage policies of this character, and where this is done the wage rate is usually expressed in terms of the standard of living.

It must be admitted that individual employers are sometimes at liberty to set their own wage rates above the point elsewhere determined by economic law. Prosperous concerns operating on a large scale, employers who possess unusual ability as managers, and above all those enterprisers who have a monopoly or something approaching monopoly in their respective lines of industry may maintain the marginal productivity of the class of workers they employ above the general level of this class in the business world. This makes it possible for them to pay wage rates above the normal or average, and certain considerations sometimes make it desirable for them to do so. Philanthropic and compassionate motives undoubtedly have some influence. Of much more importance are the considerations which make it profitable for employers with especially prosperous or quasi-monopolistic enterprises to adopt this line of wage policy.

Most employers who fix their wage rates above the normal earn-

ings of the group in question, equating them to some hypothetical standard of living, will frankly acknowledge that they do so because it is "good business." There are several possible reasons why the employer may find such a policy profitable to himself. For one thing, through his reputation of liberality he may obtain a picked lot of workers. The policy also makes for permanence of employment, reducing what is called the "labor turnover"; *i.e.*, the habit of workers of drifting in and out of jobs, often in response to whim or caprice. Since there is some expense connected with the hiring of new workers and still more involved in breaking them in to the routine of the plant, the stability of a labor force contributes to the profits of the enterprise. A contented and stable group of seasoned employees also makes for discipline in the shop and for the smooth functioning of the plant. These are but suggestions of the ways in which an attractive wage policy may be "good business." It is to be noted that these results do not follow from the fact that wages are high, but from the fact that they are higher in one enterprise than in industry generally. The industries offering the high wages do not pay their workmen more than the value of their marginal product, but they employ devices which raise this marginal value above the ordinary level of the business world. This then is no real qualification of the law of wages.

Of late the doctrine that high wages are "good business" has called forth a somewhat different line of argument in its support. It is held that our modern system of mass production presupposes a widely dispersed purchasing power in the hands of the great mass of the people and that, unless the wage-earning population possesses ample money incomes, they will be unable to take off the market the product of our highly organized industries. From this maladjustment of supply to demand will ensue business stagnation, with all its resulting injuries to society as a whole. This argument is advanced to support a policy of arbitrarily increasing the wage rates in all industries for the benefit of every social class — including the owners of business enterprise — whose welfare is involved in the smooth functioning of the productive system. Many large employers have subscribed to this proposal, Henry Ford being one of its outstanding advocates. In times of business depression the same

proposal to increase wage rates is advanced as a solution of the difficulty.

These ideas have very limited validity. Artificial increases of wages have the effect of raising business costs, and the higher costs will result in higher prices. Since the object of the policy is to expand the buying power of the general public, these effects upon prices will tend to defeat the purpose in view. It is quite possible of course that some groups of workers are at the moment being paid less than their marginal worth and that the arbitrary increase of wages will correct this condition. But as a broad social policy designed to produce general prosperity this method of artificial redistribution of income is defective in that it involves a general increase in the costs of production.

Minimum wage laws: General summary. Belief that wages too low to provide a decent standard of living are immoral or socially inexpedient has led governments to pass laws prescribing minima below which wage rates must not fall. Beginning with the legislation of New Zealand in 1894, this movement has spread through many countries until today sixteen of the principal industrial nations — among them Great Britain, France, Germany, Austria, and the British Dominions — have statutes of this character.

In the United States this form of social legislation has been in the hands of the several states except for the brief period 1933–1935, when the wage codes of the national industrial recovery act were in effect. A Supreme Court decision in May, 1935, which held this action of the federal government unconstitutional, again gave the states sole jurisdiction over the matter. After ruling variously on the constitutionality of state minimum wage laws at different times, the Supreme Court declared such laws valid in March, 1937. The spread of state laws, which had previously been retarded by doubt as to their constitutionality, began again. In 1938 twenty-four states had adopted minimum wage laws, applicable (except in one state) to women and children only.

The United States fair labor standards act of 1938. The latest attempt of the United States federal government to establish minimum wages on a nation-wide scale took the form of a law bearing the above title, which was approved on June 25, 1938. The law ap-

plies only to industries engaged in interstate commerce and within this broad classification exempts many specific occupations from its provisions. For example seamen, fishermen, agricultural workers, workers engaged in processing milk and other perishable foods, and street-car and bus-line operators are exempt. In the industries to which it is applicable the law attempts to establish maximum weekly hours as well as minimum hourly wages. It also prohibits the employment of children under sixteen years of age in these industries and sets the minimum age at eighteen years if the industry is classified as hazardous. The provisions for maximum hours are as follows: 44 hours per week during the first year; *i.e.*, until November, 1939; 42 hours per week during the second year; 40 hours per week thereafter.

The basic minimum wages set by the fair labor standards act are 25 cents per hour during the first year of its operation, 30 cents per hour during the next six years, and 40 cents per hour thereafter. An essential administrative feature of the law however introduces a large degree of flexibility into these standards, with reference not only to the dates at which the minimum wages become effective but also to the applicability of the rates themselves to certain occupational and regional groups. The enforcement of the act is entrusted to a new officer in the Department of Labor, who is empowered to appoint advisory committees, in each branch of industry, to collaborate with him in adjusting the provisions of the law to the conditions of the business and region. These committees are charged with the task of hastening the adoption of the 40 cent minimum wage over as wide an area as possible. It is a part of their duty however to classify occupational groups within each industry in order to recommend differentials from the basic rate if conditions warrant them. When so advised the administrator may alter the terms of the act in specific cases, prescribing minimum wages between a lower limit of 30 cents per hour and an upper limit of 40 cents. In sum therefore the act contemplates a schedule of minimum wages bounded by these limits rather than a universal standard. To enforce the act the administrator is given broad powers of investigation and is supported by provisions which punish offenders by fine and imprisonment.

The fair labor standards act went into effect on October 24, 1938. On that date payment of wages below 25 cents per hour by the non-exempt industries was prohibited. It is quite impossible on the basis of the brief experience subsequent to that date to appraise the effectiveness of this act or to judge of its reactions upon American industry and labor.

Appraisal of minimum wage legislation. There has been much misunderstanding of the relation of these wage-fixing policies to principles of economics. On the one hand, it is sometimes said that economic law proves the futility of minimum wage legislation; on the other, that the successful operation of minimum wage policies disproves the validity of the economic law of wages. Neither of these positions is valid. The principle that marginal productivity tends to determine competitive wage rates does not deny that wages can be raised by statute, provided the statute can be enforced. What this principle does imply is that to raise wages to an artificial level by law will create a problem of unemployment. If the legal rates are set above the marginal worth of the workers, the employer will avoid the loss involved in this situation by reducing the number of workers employed. The reduction of the labor force will raise the value of the marginal man's product to equivalence with the wage rate. This reaction to mandatory non-economic wages must always occur in an industrial system of free enterprise where employment is voluntary on the part of the employer.

This is not to say that legal minimum wages have never turned out to the permanent benefit of laborers. Indeed a study of the history of this line of public policy proves the contrary to have been the case. It is clearly shown on the record that the minimum wage laws of England have brought about a considerable improvement in the real income of certain labor groups without causing unemployment, and this is also found to be the case with the American state laws relating to the labor of women. These facts of experience argue either or both of two things. First it may be that the competitive wage in operation before the passage of the law was lower than the economic wage or, in other words, that the bargaining weakness of the workers had been exploited to the end that they were paid less than they were worth. In such a case, as we have

already seen, the actual wage rate can be raised by any device which strengthens bargaining power. Secondly workers of low productivity may have shown an increase in efficiency in response to better living conditions and the greater contentment which followed upon a rise in their standard of life. This factor too has been discussed in the preceding pages. The initial rise of wages which produces this improvement in productive efficiency may come through legal enactment, but if the ultimate result is a rise in the marginal productivity of the group, it is evident that the higher wage rate may become permanent, and that with no reduction in the numbers employed.

Whether the minimum wage will prove a real boon to labor or an ultimate evil depends then upon the caution with which the policy is applied and the extent of improvement which is attempted through it. Examination of the laws of the American states and of England which have succeeded in raising wages without creating unemployment shows that these limitations have generally been carefully observed. The usual method of determining the legal minimum has been to take the wage policy of the more generous employers in a given occupational group as a standard. The statute in most cases merely makes this policy general throughout the industry, enforcing it against those employers who have been prone to use their bargaining power to exploit the workers. Where the higher wage policy has been supported by the fact that they have had a better type of labor than the average of the group, the minimum wage law has also tended to bring the average up to this level of productivity through its effect upon the efficiency of the workers. Such experiences are quite in harmony with the economic law of wages and leave unchanged the implication of that law; namely, that any attempt through artificial means greatly to increase the wage rate of any labor group will rebound upon the workers in the form of the more serious evil of unemployment.

XX

XX

PROFITS

The word which stands as the title of this chapter is one of the commonplace expressions of everyday speech, yet it is rarely given a fixed or definite meaning. Like many other technical terms of economics, it requires careful definition in order to avoid confusion of thought. As used in economics, the term denotes a type of income distinct from the other types which have been considered in the preceding study of functional distribution; *i.e.*, a share of the social income separate from rent, wages, and interest. It is our purpose now to identify and define this particular category of income and to consider the forces which account for its existence and determine its amount.

The accountant's profit statement. As a point of departure let us place ourselves in the position of a typical business man recording the results of his year's activities. Let us suppose he is a retail coal dealer; we further assume that from other owners he leases the land and borrows all the capital employed in his business and that, instead of managing the business himself, he employs a manager. The following account, agreeing with the form with which the reader is familiar,¹ records in round numbers the income and outgo of the business for a certain year.

PROFIT AND LOSS STATEMENT, 1936

Net sales		\$100,000
Cost of goods sold		75,000
Gross profit		<u>\$ 25,000</u>
Selling and administrative expense		20,000
Net operating profit		<u>\$ 5,000</u>
Other income (none)		
Other expense		
Rent	\$1,000	
Interest	<u>2,000</u>	\$ 3,000
Net profit		<u><u>\$ 2,000</u></u>

¹ See Chapter V.

This account will suffice to indicate the essential nature of profits. It will be noted that the accountant uses the term "profit," with appropriate qualifying terms, in at least three senses. The meaning of "gross profit" is obvious. The net results of the business as an operating entity are shown by the item "net operating profit." But against this must be charged the fact that this business required that certain land and capital be devoted exclusively to its uses, thereby being withdrawn from other possible remunerative employment. The use of these factors could not have been obtained except by agreeing to pay their owners economic rent and interest respectively. And it is only after deducting these payments that we arrive at the true net earnings which go to the owner of the business as such. The accountant calls this *net profit*, and this is true *profit* in the economic sense.

Implicit rent, interest, and wages of management. Let us now assume another retail coal business, precisely like the one we have considered except that the land and capital belong to the owner and the owner manages the business himself, adding the further assumption that the manager's salary in the first enterprise was \$5,000. In our second case, — there being no item for manager's salary — the "selling and administrative expense" will be \$15,000 instead of \$20,000, and the "net operating profit" will be \$10,000. This will apparently be the final "net profit," since there are no further deductions for rent and interest paid.

In harmony with common usage, which assumes that profits are what is left to the owner out of his business after deduction of payments for costs, the ordinary "man on the street" might regard this \$10,000 as the "profits" of the year's business. But in our search for the true profit, we need to remember that the land site owned by this coal dealer could have been rented, had he so chosen, to another enterpriser for \$1,000 a year, and that the capital funds, which he had temporarily tied up in buildings and other capital instruments (trucks, shovels, coal chutes, etc.) and a stock of coal, might have been invested so as to yield him interest to the amount of \$2,000 per year. Let us also recognize that the coal dealer himself, instead of devoting his time to this particular enterprise, might have hired himself out to another coal company at an annual salary,

let us say, of \$5,000. In view of these considerations the economist insists that this particular enterprise has not really brought in a profit of \$10,000.

From this "net operating profit" of \$10,000 there must be deducted the rent, interest, and owner's salary, all three of which could have been obtained by the owner had he not gone into the coal business for himself. Only the remainder, \$2,000, can be said to have come to him on account of the business. This is the true profit. We thus reach the general conclusion that in economic analysis and in all accurate accounting, before true profits emerge due allowance must be made, not only for explicit costs (payments to non-owners), but also for the implicit rent, interest, and wages of management where these costs are not represented by payments to others.¹

Definition of profits. We have now arrived at a clear and exact concept of profits and are in position to formulate a definition. *Profits are the net income of a business, or the difference between the income and the costs, the costs including rent, interest, and wages of management whether explicit or implicit.* We must note that profits are essentially variable and uncertain. In any given enterprise they may be large one year and small the next. Costs may exactly equal income, in which case profits are zero, or may even exceed income, resulting in losses instead of profits. We shall find it an aid to our analysis to regard losses as negative profits, which makes it possible to say that profits are always obtained by subtracting costs from income, even though the former be the greater and the result therefore a negative quantity.

Actual profits. Of the actual existence of profits, as a type of social income, there is evidence aplenty. It is true that there may be difficulty in measuring profits or even in demonstrating their

¹ If the reader finds difficulty in reconciling the present inclusion of rent among business costs with our previous conclusion (arrived at in Chapter XVI) that economic rent is not a cost of production, he will do well to refer once more to the discussion at that point. From the point of view of society as a whole, economic rent is no cost, since it is a purely differential gain which does not affect the price of the product. To the individual business man, whether owner or tenant of the land used in his business, the economic rent must be regarded as one of the costs of the particular business since it would accrue from other uses of the land if the enterprise were abandoned. It still remains true, as previously pointed out, that economic rent does not affect the price of the business man's product.

existence in those cases where the business man's income includes implicit rent, interest, or wages of management. Placing a money value upon these items of business cost — particularly the last — is an act of judgment, which of necessity gives some margin of uncertainty as to the amount of the residual quantity, profits. But this difficulty is not always present. Extension of the principle of division of labor has had the effect of disclosing the pure profit income, in specific cases, in contrast with all other types of income. Many a corporation not only makes a separate accounting of rent, interest, wages, and other business costs, but hires at contractual wages all the managerial labor it requires. From the president down, all the managers are salaried employees receiving wages, which differ only in amount, not in essential character, from the wages of the humblest laborer. This practice places wages of management in the category of explicit business costs and makes it possible to earmark profits, if any exist. The United States Steel Corporation for example has at its head a hired laborer who supervises the function of management with the aid of various subordinates, who are likewise hired laborers. Each of our great railroad systems has its salaried president, vice-presidents, and department heads. So it is with the larger banks, trading companies, and manufacturing concerns.

In such cases any residuum of net income appears clearly in the light of pure profit, provided always that a proper deduction is made for interest on the capital invested in the business by the stockholders. If we focus attention on specific industries or specific individuals, we find abundant proof of the existence of this income. During the twenty-year period between 1901 and 1922 the common stockholders of the United States Steel Corporation accumulated profits of over three-quarters of a billion dollars quite in addition to dividends which they received. In one ten-year period the Standard Oil Company of New Jersey paid annual dividends of 20 per cent to its common stockholders and accumulated several hundred millions in profits. Henry Ford began his manufacture of motor cars with an investment of \$49,000 and within twenty years had reinvested profits amounting to nearly half a billion dollars. The great fortunes of modern times are based on profits.

Losses. Such examples as these however should not be allowed to stand alone. The other side of the picture discloses an impressive record of losses or negative profits.

Statistics show that at the end of each year more than one hundred thousand corporations have nothing left for dividends after payment of business costs. In 1917, a year of exceptional business prosperity, 119,000 showed an aggregate deficit of 629 million dollars. The entire business world is subject to startling reversals of fortune. In 1919 only about two-thirds, and in 1921 less than half, of American corporations reported net earnings to the United States Treasury. In 1921, out of a total of 356,897 corporations reporting to the Treasury, 185,158 showed a combined loss of four billion dollars, an amount only a little less than the combined net earnings of successful corporations. In 1932 the combined statements of 23,846 manufacturing corporations reporting to the Treasury showed that losses exceeded profits by one and a half billion dollars. These were bad years, to be sure, but they illustrate the fluctuation of earnings resulting from business risk.

The following table shows the numbers of manufacturing corporations and the percentages of the total number which reported no net income during a recent eleven-year period.

MANUFACTURING CORPORATIONS REPORTING NO NET INCOME ¹

<i>Year</i>	<i>Number of corporations reporting no net income</i>	<i>Percentage of all corporations</i>
1922	33,788	41.0
1923	31,404	36.9
1924	35,461	40.9
1925	34,537	38.9
1926	38,150	40.9
1927	36,196	40.3
1928	36,566	39.9
1929	36,742	39.8
1930	50,863	55.6
1931	58,815	66.0
1932	68,644	83.0

¹ It should be noted that this table refers to net income, not to profits. Assuming reliable accounting, it is obvious that concerns with no net income have nothing to distribute as profits.

It will be seen that on the average, even in years of prosperity approximately 40 per cent of these enterprises had no net income after the payment of business costs.

The violent fluctuation of business profit between good times and bad is shown in a recent study of the ratio of net profits to capitalization in all manufacturing corporations in the United States. In this table capitalization includes only the stockholder's equity, exclusive of bonded indebtedness.

MANUFACTURING CORPORATIONS ¹
RATIO NET PROFIT TO CAPITALIZATION
(1927-1932)

1927	7.1
1928	8.6
1929	9.2
1930	2.7
1931	1.1
1932	4.7

To avoid any possible misunderstanding, attention is called to the fact that the percentages above refer to "net profits" in the accountant's sense and not in the sense in which the term profit is employed in economics. The above percentages represent the earnings from which dividends could be paid to the stockholders and obviously include interest on the stockholders' capital as well as profits. If we assume three per cent as a reasonable rate of pure interest, it is evident that in the years 1930 and 1931 these corporations made no profits and did not even earn full interest on their stockholders' investment. Of course this qualification does not at all diminish the significance of these figures as evidence of the fluctuation of profits.

The balance of profit and loss. Our statistical data are not sufficient to permit of anything approaching an exact balancing of the total profit income against total losses. But we may at least hazard the guess that on the whole business losses are much more nearly equal to business profits, and that the net income accruing to business men as a social group is much smaller, than is popularly assumed. When visualizing the result of chance in any hazardous undertaking, it is a common failing of human nature "to count the

¹ Taken from H. G. Moulton, *Income and Economic Depression*, 1935, p. 149.

hits and disregard the misses." The casual observer of the results of economic activity in society is so blinded by the glitter of the large prizes that he overlooks the frequency of losses. But for every case of large profit in a given enterprise or for a single individual business man, a case of loss can be cited for another enterprise or individual which tends to offset profit in our accounting of the distribution of society's income. It is at least a reasonable inference that the amount of positive profit for the entrepreneur group as a whole is only a small share of the total social income.

The risk-bearing function: Business risks. True profit can exist only when there exists a total business income, as determined by sale prices, in excess of production costs. Under competitive conditions this excess tends to be eliminated by the forces of demand and supply. In strict theory, save for monopolized goods which form a class by themselves, market prices of goods and services tend to equate with normal or "bulk-line" unit costs of production, leaving no margin for pure profit. To account then for this type of income we must look to causes which bring about unforeseen changes in the conditions of supply or demand for goods offered for sale and so prevent or interrupt the adjustment of prices and costs in the market. The term *risk* may be used as a generic term to denote such uncertainties of economic life. We may classify these causes of maladjustment roughly in two general groups: those which occur as commonplace incidents in routine business operations, and those which have their roots in the dynamic character of our industrial structure. The former affect the profitability of individual enterprises within an industrial field; the latter exert their effects upon industries in their entirety.

The most striking cases of pure profit do not arise from the ordinary risks of business. Such risks, in so far as they become foreknown, lose their uncertainty and can be handled as costs. The development of the practice of insurance makes this clear. One by one many of the risks of modern industrial life have been brought within the compass of mathematical formulae and have had the law of averages applied to them. When it is discovered that a given risk is insurable, the losses which it entails are apportioned among all affected by it. It becomes one of the normal and calculable costs of

doing business and plays its part in determining market prices. Thus the business of insurance may be viewed as a continuous encroachment on the domain of the risk bearer and upon the profit income. If all hazards could be handled by the insurance method, most of the day by day maladjustments in the price system would disappear, with corresponding disappearance of the pure profit income. But heretofore it has proved impossible entirely to eliminate risk even in the routine operations of the most firmly established enterprises. The effect of the risk element is to be seen in the fluctuating character of the residual or profit income. The individual or group of individuals whose share of the proceeds from the enterprise consists in this fluctuating residuum of income perform the function of risk bearing.

But the more far-reaching of economic fluctuations are those which arise from the dynamic character of our social life, causing over a period of years changes in the processes of industry, in the range of human wants, or in the availability of natural productive agents. These changes have great power to affect the economic welfare of humanity; yet the vast majority of us take no thought about them, assuming that because we have always found the essentials and conveniences of life ready to hand in the market we may confidently expect to find them so tomorrow.

Risk bearing. Just as long as our economic system is subject to continuous change, business risks of the two types we have described will continue. As long as we rely upon individual initiative to guide our productive system, we must appeal to self-interest to induce enterprisers to bear risk and to exercise foresight in reducing risk. The work both of bearing the routine hazards of business and of anticipating future needs, discounting future hazards, and making timely adjustment to forthcoming alterations in the conditions of economic life is left to a self-constituted group of men, whom we may call the risk bearers, if we are careful to keep in mind the broad implications of the term. Their major social function is that of guiding and directing the course of economic development so that future conditions will find us prepared in advance to grapple with them. In other words, we may say that *they capitalize their prevision of the future.*

Not all business managers are equally endowed with the qualities of foresight, energy, and courage which are requisite to coping successfully with the hazards of industry. It is a natural result that among individual enterprisers there is wide variation in the amount of profits, far wider than if profits were simply the result of chance. Obviously luck is an important factor in the determination of profits, but it should be recognized that the amount of any individual's profit or loss is also the result of the degree of managerial skill and judgment with which he performs his entrepreneurial function of directing the course and bearing the risks of business.

Profit is an income similar to economic rent in that it accrues to the owners of the business. When corporations hire their managerial labor at contractual wages, the pure profit, if any exists at the end of the accounting period, is merged along with implicit interest on the stockholders' equity in a fund called "undivided profits," destined either to be distributed among the stockholders or to be reinvested in the enterprise and so merged in the value of the stock. When the enterprise is managed by the owner, wages of management remunerate him for his executive and administrative functions, and the profit, if any exists, is an income derived from his status as proprietor.

The promoters and their profits. In the case of well-established enterprises it is often difficult at first glance to discover that any productive function is being performed by the recipients of the profit income. The stockholders of a corporation are often completely ignorant of the business from which they receive dividends and have done nothing in its service save to furnish it capital, for which service they will, if the business is sufficiently successful, receive interest in addition to profits. The absentee owner of an individual enterprise and the inactive member of a partnership have legal title to their respective shares of any income in excess of business costs which may arise from the efforts of those actively engaged in the business.

Consider for example the function of underwriting new ventures as discharged by the investment banks. When they believe that timely provision of capital equipment along novel lines promises lucrative returns in the future, these banks guarantee to provide.

the necessary supplies of present funds, partly out of their own resources, partly by inducing others to take shares in the venture. Through their efforts capital instruments are constructed and assembled, management and manual labor are organized, the different parts of a productive establishment are brought into coöperation with each other. If the forecast is correct, this productive unit will be equipped and ready for service at the time it is needed. The bankers may then withdraw and surrender the actual ownership and management of the new enterprise to others. But they will have capitalized the opportunity for profit by retaining for themselves shares of stock in the enterprise, whose dividends absorb a large part of the pure profit which accrues. It often happens that these first comers in the field are compelled to divide their title to the pure profit with other investors upon whom they rely for a part of the required funds. This they may do by offering a rate of return on preferred stock in excess of the normal rate of interest, or by selling common stock at a nominal value, or even by giving it outright as a bonus to the men who venture their funds in the purchase of bonds or other securities. But a large slice they retain for themselves in the form of common stock, for which they have made no payment other than the labor of forecasting, planning, balancing possibilities of future income against future costs. Such stock is merely legal title to pure profit if it accrues. The investment bankers may retain it permanently, thus receiving a continuous stream of profit from a successful venture, or they may sell it to others after the profitableness of the venture has been established and thus acquire outright the capital value of their share of future profits.

We have just described a process known in the business world as *promoting* new industries. The class of professional promoters has been called into existence by the dynamic character of modern industry. They are a group of shrewd and farsighted business men, who devote their whole energy to forecasting the currents of change at work and placing themselves in position to profit by an excess of future income from the sale of goods and services over the expenditures entailed in costs of production. A new railroad is built across a stretch of sparsely settled prairie at a cost of scores of millions. Its promoters are predicting a thickening of population in the

region, a shifting of the geographical location of industry, a new relation between producing and consuming markets. If these changes occur, the promoters will have rendered society a service by providing capital instruments of immense worth in the new setting, and they will profit accordingly. Another foresees that the need of a great nation for meat products cannot long be satisfied piecemeal from local sources of supply. He constructs a plant for the curing and preserving of meats, perfects the technical processes, elaborates the buying and selling mechanism, and profits with the growth of the conditions which he has foreseen. Illustrations of the promoter's activity could be multiplied indefinitely. In many cases those who first foresee the opportunity capitalize their foresight and withdraw from active association with the industry, carrying with them legal title to a large share of all future profits. In other cases they remain as active managers in the new enterprise and receive wages of management in addition to profits.

The small-scale risk bearers. But we must not allow the glitter of these great prizes to obscure the fact that the service of which we have been speaking and the profit which is its recompense are involved in the most commonplace and picayune of business operations. The little corner store which appears in the midst of a straggling suburban settlement signifies someone's faith in his foresight; it may be wiped out by a shifting of industries, or a change in transportation; it may flourish on an increasing need for the services of a middleman in the little community. The tailor shop with its small fund of capital, the little factory in which are invested the savings of a group of wage earners, the grist mill on the site of a waterfall, the brokerage offices, dealers in real estate, warehouses, theatres, and hotels are alike affected by the hazards involved in the futurity of income and cost.

There is little realization of the great number of very small business enterprises in this country. The United States Department of Commerce found in 1935 that in the so-called "service industries" alone there were nearly 575,000 separate establishments, of which 429,000 did an annual business of \$2,000 or less. So it is in many other lines of production. Each of these small business men is involved to some degree in the system of risk and profit. Every

independent enterpriser, large or small, who embarks on a new line of productive service or adds to the existing facilities in the older lines is to some extent relying on his prevision of future events. Success will bring him excess of income over and above the necessary payments for his labor of management; failure will wipe out this excess of income and may sweep away both the reward for his labor and his entire investment of capital.

Risk bearing not always a productive service. This brief discussion of the source of pure profit is by no means intended as an explanation, much less as a justification, of all income to which the word profit is ordinarily applied. The forecasting of future events and the bearing of the risk involved in them are indispensable and productive services. But this is not to say that all risk is an inevitable outcome of social forces and that risk bearing as such is always productive. Risk is sometimes deliberately created so that men may profit by it at the expense of others. The gambler, including the ignorant player of the stock market, certainly takes a risk, but he does not thereby serve society. The man who "rigs" the market or spreads false reports about economic conditions multiplies the risk of other producers and may draw an income from their losses. This income he will call profit, but it is clearly a different case from the income of the man whose foresight prepares him to serve the future wants of society or to reduce its future costs. Mere foresight and the willingness to take a chance are not always a social service. The monopolist uses foresight for the purpose of exploiting the necessities of men by contributing to the scarcity of goods and services. The man who preëmpts a choice section of land is exercising foresight, but he has done nothing to ease our struggle for existence. Investment bankers have on occasion engaged in promotion activities for the purpose merely of creating an issue of securities for sale to the public without regard for the economic soundness or the social usefulness of the enterprise whose name the securities bear. In doing this they may display shrewdness above the average, but their own enrichment is no evidence that they have rendered a productive social service. If we disregard these cases of unproductive income parading under the name of profit and of true profit resulting from anti-social activities, it is not through

ignorance of their existence and of the evil they may cause. Our purpose is to emphasize the normal and the essential. It must be evident that the need for prevision of the future and of risk bearing is inherent in the very nature of our economic order and that profits are the income resulting, to those who are successful, from these services.

Profits tend to be short-lived. Pure profit, in each individual case, tends to be a temporary income. The forces of competition work incessantly toward a readjustment of prices which will wipe out the margin of income above costs. Chance gains resulting from lucky ventures or the caprice of fortune are by nature limited in duration to the continuance of the special events which have produced them. Special devices or methods of management, by which individual business men avoid loss from routine business hazards, become known to the trade in general and lead to an adjustment of prices. Even the larger flows of profits resulting from far-reaching changes in industry are of limited duration. A new invention will bestow profits on its first possessor by reducing his costs below those of his competitors. But when it becomes the common property of the trade, all costs will fall and the maladjustment will disappear. The long-run effect of improvements in machine equipment and in processes of manufacture is not to expand the profit income but to increase the real incomes of all groups in society. Even the fore-runners in new fields of industry, who are the chief profit makers in the business world, are not long exempt from competition. Those first comers may withdraw with large amounts of capitalized profits before the levelling tendency of competition has brought prices and costs into adjustment, but the industry as a whole soon ceases to be a source of profit to those who exploit it.

Profits not a price determinant. Pure profit is not one of the price determining costs of production. This income may not exist as a positive quantity for industry as a whole over a term of years. The service of the profit maker is induced by the chance of receiving the reward, not by the certainty of it. In any hazardous undertaking it is the anticipation of the reward which induces the effort of the competitors. In athletic contests, a winner's prize need be given to but one of the contestants. It is doubtless true that the service for

which pure profit is the reward would not be rendered if none of the enterprisers were able to make a profit. But by holding forth chances of gain to individuals, society is supplied with the risk-bearing service of the enterprisers as a class without guaranteeing a reward for this service in any particular case. Recalling again that we speak only of competitive industry, the following statement, analogous to the case of economic rent, can be made: "Prices are not high because profits exist; but profits exist because for other reasons sale prices are sometimes higher than costs."

Criticism of the profit system. Yet it is probably true that no feature of the capitalistic order of society has been subjected to more aggressive criticism from the advocates of social reconstruction than the system of profit industry. The more superficial of these criticisms are aimed at the size of the profit income. It is carelessly assumed that the profit income, both as a whole and in individual cases, is so large as to create great and unjust inequalities in the economic conditions of the different social classes. That this income is large in certain individual cases, we have had occasion to note; and it may be true that the social consequences of even a few such cases present a serious problem. But the customary indictment of the profit income on the score of its size fails to take adequate account of the burden of losses borne by the same social class which is the recipient of profit. We have earlier in this chapter given attention to the balance of profits and losses.

The more weighty criticisms of the profit system however are concerned, not with the income, but with the prerogatives and powers of the profit-getting class. The consequence of largest social importance resulting from the association of profits with ownership is that control over business enterprise lies in the hands of the owners and not in the hands of the managers or the workers or even the consumers, whose economic welfare is dependent upon its functioning. This allocation of power is inevitable in the present setting. The risk bearer must be the commanding officer of business enterprise, else he is powerless to perform his function. Foresight divorced from power can have no practical influence over the adjustment of the industrial structure to changing conditions. Nevertheless the arrangement is open to attack. The owner is not the

only risk bearer in a typical industrial establishment. The wage earners in particular are subjected to severe hazards from the uncertainties of business enterprise; they stand in the case of failure to lose their income through the loss of their jobs. The association of control with ownership draws a line of division between the profit getter and the worker which places the latter under the direction of the former and, without giving him a share in control, renders him liable to injury through mistakes in judgment on the part of the owner or his representatives.

These attacks upon the profit institution are not properly a part of our study at this point. They are mentioned here in order to give a glimpse of certain broad social implications of the type of income considered in the present chapter. In a later chapter we shall examine in some detail both the criticisms of the profit income and the programs of reconstruction to which these criticisms have given rise.

EXERCISES

1. Mr. X formerly was employed as department head of a manufacturing enterprise at a salary of \$6,000 per year. He had \$20,000 of savings invested in bonds which yielded him 5 per cent interest. He resigned his position and set up an enterprise of his own, investing his savings in the new enterprise. At the end of the first year his total net income amounted to \$6,200.

- (a) What were his pure profits? Explain.
- (b) If his total net income continued at this rate for ten years, would it be worthwhile for him to continue in this enterprise? Explain.
- (c) If this rate of return were typical of earnings of the industry as a whole, what would be the trend of future development in the industry? Why?

2. Suppose that Mr. X in the following year receives a total net income from the enterprise of \$10,000.

- (a) What now are his pure profits? Explain.
- (b) If this rate of return were typical of earnings of the industry as a whole, what would be the trend of future development in the industry? Why?

3. A new type of machine costs \$1,000 and yields an annual net return of \$120 on the investment. If the prevailing rate of interest is 5 per cent, what amount of pure profits will result from the use of the machine? Under conditions of competition, will the income from the new machine tend ultimately to be equalized with the income from other machinery? How?

[illegible]

Functional shares in distribution. In the preceding chapters the problem of distribution has been regarded from the functional point of view. This has involved a classification of the total income fund of society into distinct categories or classes, each category corresponding to some peculiar productive function performed by the recipient of the income or by the wealth which the recipient owns. We may now summarize the result of this analysis. The broadest classification of income from the functional viewpoint presents two major types of income: income from *ownership* and income from the *labor* or *service* of the recipient. Each of these chief groups contains a number of sub-groups as set forth in the following summary :

Sharers

Name of income

- | | |
|-------------------------------------|----------------------------|
| I. Owners of productive instruments | |
| 1. Landowners | Economic rent |
| 2. Owners of capital | Interest |
| II. Performers of human service | |
| 1. Laborers (in the broad sense) | Wages (in the broad sense) |
| (1) Employees | Wages and salaries |
| (2) Owner-managers (as workers) | Wages of management |
| (3) Professional workers | Fees |
| 2. Enterprisers (as such) | Profits |

467

TOTAL AND PER CAPITA INCOME OF THE UNITED STATES
(1799-1938)¹

<i>Year</i>	<i>Total income (Millions)</i>	<i>Total income at 1926 prices</i>	<i>Per capita income</i>	<i>Per capita income at 1926 prices</i>
1799	\$ 677	\$ 1,092	\$131	\$211
1809	915	1,423	130	202
1819	876	1,576	93	168
1829	975	2,083	78	166
1839	1,631	3,282	98	197
1849	2,420	5,450	107	241
1859	4,311	9,212	140	300
1869	6,827	8,843	180	233
1879	7,227	15,442	147	315
1889	10,701	23,780	173	385
1899	15,364	34,142	205	456
1909	26,456	48,102	292	530
1910	28,166	49,675	305	538
1911	28,104	50,096	300	535
1912	29,422	50,294	309	529
1913	31,450	53,761	326	557
1914	31,213	53,356	319	545
1915	32,533	54,042	327	544
1916	38,739	56,636	384	562
1917	46,376	57,043	454	558
1918	56,956	62,044	550	599
1919	62,945	62,199	599	592
1920	68,434	60,615	642	569
1921	56,689	59,485	524	550
1922	57,171	61,873	520	563
1923	65,662	68,044	589	610
1924	67,003	69,004	592	610
1925	70,051	70,474	610	614
1926	73,523	73,523	631	631
1927	73,966	73,966	626	626
1928	75,904	73,765	633	615
1929	79,498	75,929	654	625
1930	72,398	73,725	588	599
1931	60,203	68,647	485	553
1932	46,708	60,503	374	484
1933	44,713	59,301	356	472
1934	51,560	64,370	407	508
1935	56,254	66,337	441	520
1936	65,246	72,415	508	564
1937	69,419	73,693	537	570
1938*	62,286	69,130	478	531

* Preliminary estimate.

¹ R. F. Martin, *National Income in the United States 1799-1938*, National Industrial Conference Board, 1939, pp. 6 and 7.

The figures given in the second column of the table show a wide variation of aggregate national money income over the course of years. Since the beginning of the present century there has been a steady increase up to 1929, interrupted only by a slight recession in 1921-1923, then a striking decline until 1933, and an increase after that date through 1937. Per capita incomes were subject to similar fluctuations.

Since these incomes are stated in terms of money, it is necessary to take the changing price level into account when comparing the prosperity of one year with that of another. This can be done by dividing the incomes for each year by the index number of prices for that year, with results as shown in the third and fifth columns of the table. When thus translated into the terms of a stable price level, the changes in national and per capita income over the quarter century do not appear so striking, although substantial fluctuations still remain.

Needless to say, the figures for per capita money income do not represent the money receipts of any real individual. They are crude averages indicating only estimates of what each individual member of the population would have received in a given year if the national income of that year had been divided equally. Nevertheless these abstract estimates are of real importance in our study of personal distribution. They dispose of many false notions regarding the level of individual welfare which can be obtained by merely altering the existing pattern of distribution. In 1937 for example each person would have received only \$537 if all had fared alike, and in other years absolute uniformity would have resulted in individual incomes of surprisingly small size. We may well bear this fact in mind when it is proposed to make everybody rich by forcible redistribution of the nation's total resources.

Comparative national incomes. It would be interesting to know the comparative status of the important industrial nations in respect of the average economic welfare of their peoples. This is a subject on which we have little reliable information, and what we have consists of statistics which make comparisons between the different countries a difficult undertaking. For what it is worth the available evidence shows that the people of the United States enjoy a

much larger per capita income than the people of other great industrial nations. This is undoubtedly true to the facts, since the abundant natural resources of this country, exploited by a relatively scarce and energetic people employing a highly refined machine technology, yield up a large average product from the economic effort of the people. But comparisons of this sort can be used only as rough indications of the economic welfare of different national groups. Statistics of income in different countries have not been corrected with reference to a common price level and hence do not indicate accurately comparative real income. Still less do they prove a higher degree of contentment and happiness in the countries having the largest average money incomes. Contentment and happiness are psychological magnitudes which are as much affected by the range of wants and the customary mode of life which we call "the standard of living" as they are by the absolute incomes of the different peoples.

Income distribution in America. The year 1926 is generally taken as a norm in studies of income distribution because of the relatively stable conditions of that time as compared with preceding and subsequent periods. The following table is based on income statistics

DISTRIBUTION OF MONEY INCOMES IN THE UNITED STATES
IN THE YEAR 1926 ¹

<i>Income class</i>	<i>Cumulative percent- age of persons</i>	<i>Cumulative percent- age of incomes</i>
Under \$1,000	2.89	.31
\$1,000-\$2,000	28.15	8.27
\$2,000-\$3,000	48.40	17.57
\$3,000-\$5,000	78.37	39.56
\$5,000-\$10,000	91.92	57.24
\$10,000-\$25,000	97.88	73.91
\$25,000-\$50,000	99.27	82.81
\$50,000-\$100,000	99.776	89.14
\$100,000-\$150,000	99.881	91.74
\$150,000-\$300,000	99.960	94.75
\$300,000-\$500,000	99.982	96.30
\$500,000-\$1,000,000	99.994	97.75
\$1,000,000-and over	100.00	100.00

¹ Treasury Department, United States Internal Revenue, *Statistics of Income from Returns of Net Income for 1926*, p. 4.

for that year. The information is taken from official income tax returns showing the actual money receipts of individuals in the United States who made declarations of their incomes to the federal government.

This table does not really tell us how much money any specific individual or group of individuals received in 1926. It would appear from the figures given that approximately half of the people (48.4 per cent) received incomes of \$3,000 or less. We must remember however that the "percentage of persons" listed in the second column refers to individuals who acted as the declarers of income. The income unit in modern society however is the family, not the individual. In many cases one income tax declarer represents an entire family among whom the reported income must be divided. On the other hand, many families possess more than one income receiver. Other studies have attempted to depict the distribution of income among the families of this country with a result as shown in the following table.

DISTRIBUTION OF INCOME BY FAMILIES ¹
(The United States 1910 and 1929)

<i>Income Class</i>	<i>Per cent of families</i>		<i>Per cent of total national income</i>	
	1910	1929	1910	1929
Richest Families	2	2	20	27
Middle-class Families	33	33	41	44
Upper Middle-class	18	18	27	29
Lower Middle-class	15	15	14	15
Poorest Families	65	65	39	29

Inequality: In America. The most striking feature of the pattern of distribution in America is the extreme inequality which it portrays. Under conditions of complete equality any 10 per cent of the people would receive 10 per cent of the national income, any 50 per cent of the people would receive half of the income, and so on. A glance at the table of individual incomes will show how far from this line of uniformity is the actual pattern. As things were in 1926,

¹ Taken from C. R. Daugherty, *Labor Problems in American Industry*, Revised Edition, 1938, p. 155.

the lowest income class included 2.89 per cent of the people and received only .31 per cent of the total income. A group of about the same size at the top of the scale (the 2.12 per cent receiving over \$25,000 a year) got 26.09 per cent of the national income. The richest 1 per cent received 18 per cent of the total income, while at the bottom of the scale a section of income similarly large was shared by nearly half the people; in other words, the poorest 48.40 per cent of the population had a share a little smaller than that of the upper 1 per cent. In whatever manner the table is analyzed, this condition of inequality is thrown into sharp relief.

The figures of family distribution not only call attention to extreme inequality but appear also to indicate that during the period under review this inequality was increasing. The richest families, 2 per cent of the total number of families, received 20 per cent of the whole national income in 1910 and 27 per cent in 1929. The middle-class families, approximately a third of the total, showed less variation, their share of the whole income increasing from 41 per cent in 1910 to 44 per cent in 1929. The 65 per cent of the families standing at the bottom of the income scale received 39 per cent in the former year, 29 per cent in the latter. Too great reliance should not be placed in these figures as evidence that the rich are growing richer and the poor poorer over long-run periods; not only is the time covered by the table too brief to show a permanent trend, but it was disturbed by war and its aftermath and phenomenal technological changes. However we are not concerned here with the developmental tendencies in modern society but rather with its general pattern of distribution at a given time.

Inequality universal. Similar conditions of inequality are to be found in all countries for which studies of personal distribution have been made. As evidence of the general type of distribution on the continent of Europe, the statistics of income taxation in Prussia are illuminating. Taking the figures for the year 1908, we find that out of a total population of about 38,000,000, 45 per cent were exempt from the income tax because their incomes were less than 900 marks (\$225 per year). Nine-tenths of the taxpayers had incomes of less than 3,000 marks; only one-tenth, incomes exceeding this amount. A similar analysis of the British income tax

returns for 1904 shows that less than 10 per cent of the people received half of the total income of the country. These figures are typical of all modern countries. Everywhere the largest number of individual incomes is massed at the lower end of the income scale, while the small number at the upper end of the scale receives the lion's share of the total income.

Perhaps it is well to point out that these figures for the distribution of money incomes do not indicate the comparative scales of consumption or enjoyment of the different individuals and social classes. The upper middle and the highest income classes consistently save and invest a portion of their money incomes; the lower income classes save very little. Saving reduces the amount of current consumable income, and continued reinvestment of the income from savings leaves the recipient in the same position, in respect of consumable income, as if his original money income had been proportionately smaller. In the case of many of the largest incomes, this process of continuous postponement of consumption is carried on throughout the lifetime of the income receiver. To what extent the practice of saving affects personal distribution of final income we do not know since we have no accurate statistics of the amounts saved by the various income classes. This factor must have the effect of mitigating to some degree the condition of inequality pictured in our statistics; but it tends to accentuate inequality in the distribution of wealth (as distinct from income) and through inheritance perpetuates inequality in succeeding generations.

Personal distribution of wealth: The evidence. If we turn from the distribution of income to that of wealth we discover a similar, but even more extreme, state of inequality. Wage earners as a class receive a large fraction of the total national income, but most of them own only trifling amounts of wealth, and their total of property rights represents but a small portion of the nation's wealth resources. In seeking statistics of the distribution of wealth we encounter even greater difficulties than in the case of income. There is of course no census of ownership to be drawn upon for information regarding this problem. The most useful sources of data are the records of the probate courts, through which estates pass upon the death of their owners. The statistics assembled in

the following paragraphs are taken from these records. It would be well before examining the data to point out certain inaccuracies necessarily resulting from this method of estimating the distribution of ownership. In the first place, an individual's accumulation of wealth is not always transmitted as a whole upon his death; many family estates are divided up in whole or in part during the owner's lifetime. This practice will cause the estates passing through the probate courts to be smaller on the average than are the property holdings of individuals still living. Secondly, it must be remembered that the average age of those who die in a given year is higher than the average of the living population, and the decedents have therefore had a longer time in which to accumulate wealth. This factor will have an effect opposite to that just mentioned; it will cause the record of the probate courts to exaggerate the size of individual estates. Because of such uncertainties these records cannot be used with safety to estimate the actual amounts of wealth owned by different social classes at any given time. Our real interest in the problem however is to depict the pattern of distribution rather than to discover the absolute size of different individual estates, and the data we are using are significant for this purpose.

In America. The following table gives a simple analysis of wealth distribution as shown by the probate records of two American states, Massachusetts and Wisconsin, on the dates given. The figures listed in the fourth column are of some interest as indicating the average amount of property held by the respective categories of the population: the poorest, lower middle, upper middle, and richest classes. However for the reasons given above they are not entirely reliable. What is more important for our purposes is a comparison of the percentages given in the first and third columns of the table, for it is this that discloses the extent of inequality in wealth ownership.

If wealth had been divided evenly in these states on the dates given, each fraction of the population would have owned an equal fraction of the wealth which passed through the probate courts. Instead of this situation we find that the poorest two-thirds of the people had a combined ownership of 4 or 5 per cent of the wealth. The lower middle class, 15 per cent of all the people, owned a some-

what smaller proportion. The richest 2 per cent held nearly three-fifths of all the wealth. Or to throw the figures into somewhat different arrangement, the distribution can be given very simply as follows: 80 per cent of the people own 10 per cent of the wealth; the remaining fifth of the people have all the rest — 90 per cent of the total.

SECTIONS OF THE POPULATION CLASSIFIED ACCORDING
TO WEALTH OWNERSHIP ¹

<i>Section of population</i>	<i>State</i>	<i>Per cent of total wealth</i>	<i>Average value of wealth</i>
Poorest 65%	Mass. 1891	4.5	\$ 399
	Wis. 1900	5.2	381
Lower Middle Next 15%	Mass. 1891	3.9	1,499
	Wis. 1900	4.8	1,524
Upper Middle Next 18%	Mass. 1891	32.8	10,509
	Wis. 1900	33.0	8,730
Richest Class Next 2%	Mass. 1891	58.8	169,550
	Wis. 1900	57.0	135,715

For the reasons given the evidence of probated wills probably exaggerates the extent of inequality in the distribution of wealth. Nevertheless the most careful studies of the problem using other data show that most of the wealth of this country is owned by a minor fraction of the people. Thus in 1927 it was found that wealth was distributed substantially as shown in the following table, whose figures have been grouped so as to make them comparable with the figures of income distribution presented earlier in this chapter.

ESTIMATED DISTRIBUTION OF WEALTH, UNITED STATES, 1927 ²

<i>Property class</i>	<i>Per cent of population</i>	<i>Per cent of total national wealth</i>
Richest Class	2	40
Middle Class	33	45
Poorest Class	65	15

¹ W. I. King, *Wealth and Income of the People of the United States*, 1915, p. 79.

² W. I. King: Wealth Distribution in the Continental United States, *Journal American Statistical Association*, June, 1927.

In other countries. Studies have also been made of the available statistics of wealth distribution in France, the United Kingdom, and Prussia. The estimated distribution of wealth in these three countries is shown in a simplified form in the table below. The figures for the United Kingdom and France are based on probate records; those for Prussia, on tax returns.

There are certain minor differences in detail shown by these figures from the different countries, but the similarities are much more striking. Everywhere the general scheme of wealth distribution is the same; a minute upper fraction of the people owns considerably more than half of the total wealth; a large major fraction of the total population at the bottom of the scale owns virtually nothing; between these two extremes we find a lower middle class with a fraction of the total wealth much less than proportionate to its numbers and an upper middle class with somewhat more than its proportionate share. All other evidence which we possess points to the same general conclusion regarding wealth distribution in nations of advanced civilization.

SECTIONS OF THE POPULATION OF PRUSSIA, FRANCE, AND THE UNITED KINGDOM CLASSIFIED ACCORDING TO OWNERSHIP OF WEALTH ¹

<i>Section of population</i>	<i>Country</i>	<i>Per cent of total wealth</i>	<i>Average value of wealth</i>
Poorest 65%	Prussia 1908	4.9	\$ 153
	France 1909	4.3	186
	U. K. 1908	1.7	133
Lower Middle Next 15%	Prussia 1908	5.5	743
	France 1909	5.6	1,052
	U. K. 1908	2.9	979
Upper Middle Next 18%	Prussia 1908	30.6	3,445
	France 1909	29.4	4,602
	U. K. 1908	23.7	6,670
Richest Class Next 2%	Prussia 1908	59.0	59,779
	France 1909	60.7	85,500
	U. K. 1908	71.7	181,610

Labor's share of the national income. In view of the great inequality which exists in the distribution of wealth, it is of some

¹ W. I. King, *National Wealth and Income*, 1926, p. 96.

importance to know what fraction of the total income accrues to the wealth-owning classes, and what fraction is distributed as a reward for labor. This is a subject on which we have no precise information at all. Individual incomes are often drawn from several different sources and represent several types of productive function. It is impossible to classify the income receivers of the country under the four headings which we have adopted for our study of functional distribution — landowners, capitalists, laborers, enterprisers. Almost the only functional distinction which we can draw between the income receivers of the industrial world is that which separates the hired workers on the one hand from all other sharers in the product of industry on the other.

The following table presents the best information available regarding the division of income between the recipients of wages and salaries, on the one hand, and all other participators in the income distribution, on the other. It should be noted that these figures refer to income paid out by American industry and do not therefore embrace all incomes received, for example the incomes of the self-employed. Also the final column includes incomes which we have classified as wages in the broad sense — wages of management. The table really indicates the division in industry between hired workers, on the one hand, and employers and property owners, on the other.

RELATIVE SHARES OF CERTAIN FUNCTIONAL
GROUPS IN NATIONAL INCOME ¹

Year	<i>Estimated percentages of income paid to</i>		
	WAGE-EARNERS	SALARIED EMPLOYEES	ENTERPRISERS AND PROPERTY OWNERS
1910	39	16	45
1918	36	20	44
1921	44	23	33
1929	42	22	36
1932	39	23	38
1935	43	21	36

The table of course does not prove that the actual incomes rose and fell in absolute amount in harmony with these percentages

¹ From C. R. Daugherty, *op. cit.*, p. 151.

since the total income to be divided varied sharply during the period. Furthermore the average wage or salary in different years would be affected by any increase in the relative number of hired employees as compared with the number of employers and property owners. What is shown is the changes in the size of the shares going to different functional groups. A significant feature is the evidence that in times of depression a larger share goes to the workers who are fortunate in holding their jobs than to the property interests.

Inter-relation of wealth and income distribution. In the foregoing pages we have considered the distribution of income and that of wealth as separate and distinct phenomena. It must be obvious however that the two things are inter-related in many ways. In the first place, unequal distribution of wealth is one of the causes of the inequality of personal incomes. Most of the large aggregations of wealth are productive of income in the sense in which we have been using the term; *i.e.*, money receipts. The small property holdings of the poor are almost invariably non-productive in this sense; they consist of household and personal effects. The distribution of income in any society would be much more nearly uniform if income-bearing property were shared equally. This point is so self-evident that it requires no elaboration. Still it may be of interest to present some definite information upon it with reference to the situation in the United States. The Bureau of Internal Revenue has published an elaborate analysis of the percentages of their total incomes drawn from property and labor respectively by various classes of the population in 1930. These facts are summarized in the following table:

PROPORTIONS OF INCOME FROM EARNINGS AND PROPERTY, RESPECTIVELY
1930

<i>Income group</i>	<i>Per cent of total income provided by earnings</i>	<i>Per cent of total income provided by property</i>
Over \$1,000,000 a year	2.2	97.8
\$25,000-\$50,000 a year	29.5	70.5
\$5,000-\$10,000 a year	62.2	37.7
\$1,000-\$2,000 a year	82.2	17.8

These figures seem to show that incomes would still be very unequal, though much less so, if income-bearing property were not privately owned. It should be borne in mind however that the portion of the rich man's income which he attributes to labor is often very arbitrarily determined. The owner of a large enterprise can assign himself a large salary, leaving proportionately less to accrue to him in the form of dividends.

Is inequality increasing? It has long been a controversial question whether the tendency of capitalistic society is toward increasing inequality of welfare among society's members. Many statements which purport to throw light on this question are beside the point. For example there can be no doubt that the level of comfort for all classes in modern capitalistic society is higher than at any previous period of history and higher also than the level of comfort in modern non-industrialized nations. The poor in the capitalist society of today enjoy satisfactions which were formerly within reach only of the wealthy. Absolute misery and destitution are on the decrease and have been decreasing throughout the whole of modern history, if we except periods of civil anarchy or international war. But this does not prove that inequality is diminishing.

In America. In the United States inequality is probably greater now than it was a century ago. Like all new countries, America began with an abundance of natural resources and a relative scarcity of labor and capital. The share of the social income allocated to ownership was small. Most incomes were drawn from labor, and most laborers were given opportunity to apply their effort to natural resources which were practically unlimited. It is the competition of life, the struggle of man with man for control over limited income-producing opportunities, which brings out differences in the inherent abilities of individuals and produces inequality of fortune. On the frontier most people lived in a condition of rude comfort on a plane of comparative equality. The passing of the frontier, the multiplication of the population, and the growth of capitalism altered this condition.

During the nineteenth century all classes increased their prosperity, but the rate of increase for the well-to-do was faster than

the rate for the poor. The appearance of capitalism in industry, with its immediate tendency toward large-scale production, had the effect of concentrating income in the hands of the possessing classes. The first comers in the field of business organization reaped large rewards in the form of profits. Rising rents and land values contributed to the income from ownership, especially in urban enterprises and in certain sections of the mining industry. At the same time floods of immigrant labor retarded the rise of wages for the mass of workers at the bottom of the scale. The outcome of these forces during the last half of the nineteenth century is to be seen in the rapid multiplication of large fortunes. For the lower and middle classes economic improvement was steady, though less rapid.

The evidence presented in the table on page 471 would seem to show that in the years preceding the depression of 1929 the tendency was to increase the prosperity of the rich faster than that of the poor. But as pointed out in the discussion, it is not safe to draw final conclusions from this brief period of time. Whether the earlier trend in this country toward increasing inequality will continue now that the economic system has attained maturity cannot be known until the country settles down again to stable conditions.

Evidence from other countries. Information on this question as regards other countries is disappointingly meagre. The classic analysis of the trend of distribution over a recent period was made by A. L. Bowley in *Change in the Distribution of the National Income, 1880-1913*, published in 1920. This study is based on the income tax returns of Great Britain. Its net result is that no significant change has occurred in the status of distribution during the period under review. All classes shared about equally in the increasing prosperity of the country; the well-to-do class received about the same share in 1913 as in 1880; the fraction going to property, as compared with that accruing to labor, was also unchanged.¹

Examination of the tax returns of Germany during the same period shows substantially the same result. The average income rose, and all classes shared in the increase. In that country, as in the United States, because of the more rapid pace of industrialization there was greater tendency toward increased inequality than

¹ A. L. Bowley, *op. cit.*, p. 22.

in England. The incomes of the wealthy classes of Germany increased at a relatively faster rate than did those of the poorer classes, but the difference was not sufficiently striking to support a conclusion that increasing inequality is an inevitable characteristic of our system of industry. What other information we possess agrees with the results of these studies.¹ In normal times and as a general rule there seems to be no evidence that capitalism in a well developed form operates to increase the riches of the wealthy and the poverty of the poor.

The problem of poverty. We have been discussing the problem of inequality, which, as has been pointed out, is not the same thing as the problem of poverty. The question of the actual welfare of the people is however an important aspect of our subject. What for example is the grand total of income available to the people of the United States? How well off would each person be if all had an equal share? How well off is the mass of people under the existing scheme of division?

The first two questions are answered by the table given on page 468. To answer the third question adequately we need three types of information: (1) the average size of families; (2) a standard of living, representing the minimum necessary for health and decency for American families, against which to measure the conditions which actually prevail; (3) reliable statistics of the actual money receipts of these families. A word of explanation should be given regarding these three points before proceeding with our inquiry.

The family unit. Division of income in our type of society is carried out by families, while statistics of income, such as wage rates, are given in terms of individuals. To appraise the adequacy of a wage earner's income we need to know how large is the family which relies on it for support. The size of the family is also an important element in our computation of the standard of living. Now in such studies as this it is customary to assume that the average family consists of five persons: mother, father, and three children. This is justified on the ground that three children per

¹ V. Pareto, *Cours d'économie politique*, Vol. 1, contains studies of the distribution of income in different countries, chiefly in the nineteenth century. These studies show a marked similarity in the scheme of distribution in different places and at different times.

family are required to maintain the level of the population, and of course social income could hardly be considered adequate if it failed to keep the population as a whole alive from generation to generation. It is also assumed that each family is supported on the earnings of one member — the breadwinner — and has no income aside from his labor. These assumptions are made in the statements which follow. They are, as a matter of fact, rather wide of the truth. Investigations have shown that 27 per cent of all adult males gainfully employed are unmarried; that two-thirds of our wage-earning families have less than five members; that half the families obtain supplementary earnings from others than the chief breadwinner; and that wage earners as a whole draw between 5 and 8 per cent of their income from property.¹ These are important qualifications, tending to mitigate considerably the condition of poverty pictured by the figures that will be given.

The “standard of living.” There is no absolute “standard of living.” When we say that a certain minimum of money income is needed to support a family we always reflect our preconceptions of what decent living should be; and these preconceptions vary from time to time and from country to country as conditions and customs change. Nevertheless it is customary to measure the actual economic condition of families against an objective standard, determined by carefully estimating the kinds and quantities of food, clothing, housing, and other items of consumption and of saving required, and then resolving these estimates into money terms at prevailing prices. As a matter of fact, several different standards are usually recognized at a given time, the most important being the *minimum-health-and-decency* standard and the *minimum-comfort* standard. The former provides respectable housing, nourishing but inexpensive food, adequate and decent clothing, and a small surplus for such items as cheap recreation, a newspaper, and the like, but no saving for emergencies. The latter provides these things somewhat more generously, particularly in the case of housing, and leaves a margin for secondary education, insurance, and medical

¹ Should this statement appear in conflict with the last line of the table on page 478, the reader need only recall that the earlier table relates to those who filed United States income tax returns, a group which represents only a very small fraction of the wage-earning class, and the wealthiest part at that.

care. Incomes slightly below the health-and-decency standard are interpreted as bare subsistence; anything less than this is recognized as outright poverty. The minimum-comfort standard corresponds roughly to what is meant by the expression, the "American" standard of living.

The costs of these standards vary from year to year with price fluctuations; they differ at any one time in different localities, and needless to say, they will differ too for families of different sizes. The costs of the different standards, estimated for a family of five in the years given, are shown in the table below. The subsistence level omitted from this table may be assumed to lie in the area between the poverty and the health-and-decency standards.

ESTIMATED COSTS OF DIFFERENT STANDARDS
OF LIVING FOR A FAMILY OF FIVE

<i>Year</i>	<i>Poverty level</i>	<i>Health and decency</i>	<i>Minimum comfort</i>
1910	\$ 570	\$1130	\$1700
1918	930	1860	2780
1929	1000	2000	3000
1932	810	1610	2420
1935	850	1700	2550

Family incomes. Actual family incomes are difficult to discover. Wage rates do not give us this information in accurate form, since total earnings are affected by unrecorded factors, such as loss of work through illness and unemployment, overtime earnings, and supplementary income from other sources. The figures we shall use are based primarily on wage rates, checked wherever possible by direct inquiries among wage-earning families regarding the actual size of their incomes. Statistics relating to this problem take account of price fluctuations over periods of time, but otherwise they necessarily deal in averages which conceal local differences in living costs and actual variations in the size of families. Accordingly the data presented below ¹ as to the distribution of American working-class families among the different planes of living are

¹ From C. R. Daugherty, *op. cit.*, pp. 138 and 144.

merely suggestive of the conditions which obtain. As averages they are subject to the errors just mentioned, and they also fail to take account of many supplementary incomes referred to in a preceding paragraph. They do not tell us precisely how many actual human beings in this country were living on these planes in the years given.

The facts. Taking but two dates in recent history, 1914 and 1920, we discover the following facts regarding the economic well-being of the mass of people in the United States. In 1914, the minimum standard of living for laborers' families in New York City, according to an estimate of the city government, cost \$845 a year. For the state as a whole another official estimate placed the minimum cost at \$876. For the country as a whole the cost, according to an independent estimate of a careful student of the problem, was \$825. We have three estimates of the actual family earnings of industrial workers at this time, one for New York State, a second for Pennsylvania, a third for the entire country. In New York State industrial and office workers could earn \$649 if fully employed. In Pennsylvania the actual earnings of wage earners averaged \$720 a year. A nationwide survey of actual incomes showed that four-fifths of the heads of families in the wage-earning population earned less than \$800 a year.

By 1920 the cost of living had risen. Estimates of incomes needed to support a minimum standard of life varied from \$1,700 for an irreducible standard of health and decency to \$2,000 for a standard of reasonable comfort. In this year it was found that factory workers in New York State could earn about \$1,500 a year if employed full-time. Throughout the country as a whole workers of all grades of skill were receiving family incomes which averaged \$1,455. The average family incomes of skilled factory workers were approximately \$1,700.¹

The dates we have taken for purposes of illustration were times of fairly normal industrial activity, although it is to be noted that estimates of living standards for 1920 are abnormally high on account of the very high price level in that year. Between 1920 and

¹ For a comprehensive study of incomes and living costs see A. Epstein, "American Wages, 1890-1920, and the Standard of Living," *Annals of the American Academy of Political and Social Science*, Sept., 1921.

1929 earnings rose somewhat more rapidly than living costs, and the working population was accordingly more prosperous. Since 1929 poverty has of course increased; but the times have been too disrupted by the business depression to serve as an index of the general welfare.

The table which follows records the distribution of all families, regardless of size, among the different planes of living, as estimated by one student of the problem.¹

CLASSIFICATION OF FAMILY INCOMES
(Percentage of All Families regardless of Size or Class)

<i>Year</i>	<i>On poverty planes</i>	<i>On subsistence planes</i>	<i>On health and decency planes</i>	<i>On comfort planes</i>
1910	15	55	20	10
1918	15	45	25	15
1929	20	35	20	25
1932	40	35	15	10

The figures given suffice to show that even in a country as rich as the United States, certainly the most prosperous of the modern world and probably more opulent than any other society in human history, poverty is the lot of the lower income classes.

Causes of inequality. It is customary to divide the causes of inequality into two groups: (1) those which inhere in the nature of man, and (2) those that arise from the nature of the institutions of society. Authorities on this subject differ in the amount of stress which they place on these two factors. There are those who trace all differences in material fortune to differences in the natural endowments of individual men. This conviction leads to two conclusions: first that existing inequalities are a fairly precise measure of differences in the economic services rendered by different individuals and social groups, secondly that the present state of distribution is determined by natural law and cannot be materially altered.

On the other hand, there are many who believe that existing

¹ C. R. Daugherty, *Labor Problems in American Industry*, Revised Edition, 1938, p. 144.

inequalities are created and perpetuated by social arrangements. The rich owe their favored position to privilege sanctioned by social force rather than to their own intrinsic merit. Those who hold this view are inclined toward reform, believing that there is no obstacle other than established custom in the way of intentional redistribution of income in the direction of greater uniformity.

In our social order difference in personal ability is one cause of inequality of income, but it is not the only one. No social order whose economic structure is founded on individual initiative and free enterprise can avoid inequality of personal incomes, as long as men begin the competition of life unequally endowed by nature. In addition the institution of inheritance acts in each generation as an independent cause of inequality and through a period of time as a cumulative force perpetuating past differences in income.

The problem of inequality. It is not necessary to undertake the impossible task of defining the "best" type of income distribution in order to see that extreme inequality is an evil. The only measure we have of the final result of society's economic effort runs in terms of utilities obtained by the ultimate consumers. We may safely infer from the general law of diminishing utility that extreme inequality of distribution reduces the aggregate of human satisfactions derived from a given amount of social income. This does not prove that flat uniformity of distribution is an ideal state of affairs from the standpoint of society's long-run interest, since this would be to reward men without regard to the economic value of their services and hence to discourage initiative and industry. But it suggests that any lessening of inequality, resulting from a free play of economic forces, will increase the benefit from society's productive effort.

Furthermore there can be no doubt that the economic welfare of a social group over a period of time is vitally affected by the stability of its institutions and the smoothness with which they function. This is determined in large degree by the contentment of the average man with his lot in the social structure. Discontent and rebellion are engendered, not so much by absolute poverty and misery, as by inequality of fortune, the relative poverty of

the majority in contrast with the opulence of the few. It is now generally admitted that any arrangements which tend to accentuate inequalities between social classes are likely to be impermanent. They develop resentment and resistance which tend ultimately to destroy them. And the process of the modification of such social institutions is likely to be accompanied by such turmoil and disorder as will tend to impoverish the whole social group.

The concern over the existing scheme of distribution has gathered such strength that all modern industrial nations have begun to experiment with various programs of reform. Since it is our purpose in later chapters to examine certain of these programs in some detail, we shall not attempt in this place to do more than summarize their general characteristics.

Reform of inequality: Charitable relief. In a group by themselves may be placed those devices which are intended, not so much to remedy inequality, as to relieve the distress of the poorest members of society. Official and private charitable relief was never so abundant in the history of the world as at the present time. This is not due entirely to an elevation of human character; in large part it is a reflection of the increasing average prosperity of the people, which throws each case of misery and destitution into clear relief. It is doubtful whether charity alone will cause any permanent improvement in the problem of poverty, and clear it is that indiscriminate charity must be reckoned among the causes of that problem. Much progress has been made in the administration of charity, both public and private, toward the end that relief of poverty may have a constructive effect in building up the self-dependence of destitute families and individuals. But society has not yet recognized what seems to be a necessary implication of a condition of dependency; namely, that the individual who is unwilling or unable to bear the cost of his own maintenance has no right to marriage or to the procreation of children after marriage. The movement toward the control of births among the hopelessly unfit is gaining headway in America, some twenty-seven states having passed laws dealing with the problem. But the proposal to exert the same control over destitute families has scarcely grown beyond the stage of academic discussion.

Social insurance. Similar in nature to the relief of distress, but of more importance in that they attack the causes and not the symptoms of this evil, are those devices which aim to forestall or eradicate the hazards of life which are so frequently the immediate cause of poverty. Sickness, unemployment, accident, and premature death are normal risks of the modern workingman's life. In the lowest income classes, where the normal wage provides no margin above necessary living costs, the occurrence of one of these disasters is often sufficient to destroy the economic basis of the family. Most of these hazards are insurable risks. When the individual subject to them cannot or will not bear the cost of insurance, it is in society's interest either that he be compelled to do so or that insurance be provided for him at public expense. Compulsory social insurance has so completely demonstrated its utility as a partial solution of the problem of poverty that it has been accepted as a normal feature of industrial life in all modern civilized nations.

Preventive medicine, public clinics, public sanitation, and health movements may be considered of the same social significance as compulsory social insurance. They are safeguards, built up at public expense and in the public interest, against the hazards of life which threaten the lowest social classes with misery and destitution.

Public education. Aside from these devices for the relief or prevention of outright poverty, the programs which aim at greater equality in distribution must work some more or less far-reaching change in the existing institutions of society. The general program for free, compulsory education is perhaps an exception to this statement. To the degree that education breaks down the barriers between labor groups, it will tend to lower the higher wage income and raise the lower, at the same time increasing the average. But we do not know to what extent universal education will have this effect of promoting transference from one labor group to another, since it has never been demonstrated that the differences in productive ability which now divide the working population into classes are not due to inborn characteristics. In any event education will not remove inequality in distribution, since it will not

equalize the capacities of individuals and will leave untouched the institution of inheritance, which at present prevents an equalization of opportunity. Thoroughgoing change in the present scheme of distribution would involve a modification of our existing institutions of individualism and private property.

Limitation of personal and property rights. The programs of the socialists and the communists involve a complete transformation of these institutions of capitalism. Leaving these programs for later discussion, we may note that in most countries today the governments have begun to consider measures looking toward bringing about a greater diffusion of income and wealth. Individual liberty and free enterprise are not uncontrolled and unlimited. Many business practices which were permissible a generation ago are now forbidden because they enrich the strong or the unscrupulous at the expense of other members of society. In certain cases governmental bodies have taken over the ownership and operation of certain forms of enterprise which were formerly a source of private profit. The powers of the monopolist are more and more circumscribed by public regulation of his prices. The weakest bargainers in the laboring classes are protected in many countries by minimum wage laws which prevent their exploitation. This many-sided movement toward restriction of individual freedom in economic affairs is tending to prevent the enrichment of individuals by methods which are condemned by prevailing standards of justice and fair play.

Similarly property rights have been modified by the movement toward greater equality of income and opportunity. Graduated income taxes, in conjunction with the expansion of governmental expenditure along lines of general public service, result in an indirect transference of income from the rich to the poor. Inheritances are subject to graduated taxes, which tend substantially to reduce the largest fortunes. All of these methods of modifying the operation of the industrial system in the direction of greater equality of income and opportunity may be illustrated from the practices of any modern government.

EXERCISES

1. Refer to the table on page 468.
 - (a) Draw, on the same diagram, a curve to indicate changes in per capita money income and a curve to indicate changes in per capita real income (*i.e.*, money income corrected for changes in the price level).
 - (b) Which shows the greater fluctuation? What was the per cent decline in per capita money income from 1929 to 1932 and what was the corresponding per cent decline in per capita real income?

SUGGESTIONS FOR FURTHER READING ON PART III,
THE DISTRIBUTION OF WEALTH AND INCOME

I. GENERAL

- HOBSON, J. A. *The Economics of Distribution*. New York, 1900
CLARK, J. B. *The Distribution of Wealth*. New York, 1900
MARSHALL, A. *Principles of Economics*. Eighth edition. London, 1920.
Books IV, V, VI

II. RENT

- WALKER, F. A. *Land and Its Rent*. Boston, 1883
DORAN, H. B., and HINMAN, A. G. *Urban Land Economics*. New York, 1928
ELY, R. T. *Outlines of Land Economics*. New York, 1924
GEORGE, H. *Progress and Poverty*. New York, 1880
TAYLOR, H. C. *Outlines of Agricultural Economics*. New York, 1931

III. INTEREST

- FISHER, I. *The Theory of Interest*. New York, 1930
MOULTON, H. G. *The Formation of Capital*. Washington, 1935
CASSELL, G. *The Nature and Necessity of Interest*. London, 1900
MILLIKAN, MAX. *Liquidity-Preference Theory of Interest*. American Economic Review. June, 1938

IV. WAGES

- DOUGLAS, P. H. *The Theory of Wages*. New York, 1934
HICKS, J. R. *The Theory of Wages*. London, 1932
HAMILTON, W. A., and MAY, S. *The Control of Wages*. New York, 1923
BRISSENDEN, P. F. *Earnings of Factory Workers, 1899-1927*. U. S. Bureau of the Census. Washington, 1929
DAUGHERTY, C. R. *Labor Problems in American Industry*. Chap. VII. Revised edition. Boston, 1938

V. PROFITS

- KNIGHT, F. H. *Risk, Uncertainty, and Profit*. New York, 1933
FOSTER, W. F., and CATCHINGS, W. *Profits*. Newton, Mass., 1925
EPSTEIN, R. C. *Industrial Profits in the United States*. New York, 1934
TUCKER, R. *Is There a Tendency for Profits to Equalize?* American Economic Review, Sept., 1937

VI. PERSONAL DISTRIBUTION

- KING, W. I. *The Wealth and Income of the People of the United States*. New York, 1915
- National Bureau of Economic Research. *Income in the United States*. New York, 1921
- KING, W. I., and EPSTEIN, R. C. *The National Income and Its Purchasing Power*. New York, 1930
- MARTIN, R. F. *National Income in the United States, 1799-1938*. National Industrial Conference Board. New York, 1939
- COPELAND, M. A. *The National Income and Its Distribution*, Chap. XII. In *Recent Economic Changes*. New York, 1928
- United States Bureau of Internal Revenue. *Statistics of Income for 1936*. Washington, 1938
- MILLS, F. C. *Economic Tendencies in the United States*. New York, 1933
- Senate Document, No. 124, 73d Cong. 2d Session. *National Income, 1929-1932*. Washington, 1934
- LEVEN, M., and WRIGHT, K. R. *Income Structure in the United States*. Washington, 1938
- KUZNETS, S. *Studies in Income and Wealth*. New York, 1937

XX

•

XXII

MONEY. GENERAL PRINCIPLES. STANDARD MONEY SYSTEMS

The modern economic structure is reared on the foundation of division of labor and exchange of products. This was the fundamental thesis of Part I of the present book. Part II was devoted to the principles of value, which regulate exchange and through exchange control the production and distribution of wealth and services. The distribution of wealth and income, treated in Part III, is brought about chiefly by means of exchange. Exchange, in its modern omnipresence and complexity, would be impossible without an efficient medium, which brings us to the subject of money, the medium of exchange. Money and its correlative, banking, furnish therefore the theme of the fourth part of this volume.

Nature of money: Examples of primitive money. Money today and in the past appears in such a variety of forms that one's first steps in its study are apt to produce a sense of bewilderment. There are coins of gold, silver, nickel, bronze, and other metals. There are pieces of paper inscribed with promises and other contracts. Cattle and sheep have been used as money, as well as wheat, corn, tobacco, musket balls, beaver skins, fish, meat, shell beads, gold dust, and so on. Monetary history discloses that each community has inclined to adopt as its money some commodity or commodities of common use among the people, something that is desired by most of the people most of the time.

One of the earliest forms of money of which we have knowledge consisted of the sheep and cattle of peoples on the pastoral stage of economic development. Most of the people derived their living from their flocks, and these represented their principal form of wealth. One would always welcome an addition to his flock, and thus it became the natural thing to use the domestic animals as money. The word "pecuniary," derived from the Latin word *pecus*

for cattle, is an interesting heritage of the early stage in which cattle and money were in a sense synonymous.

The early settlers of Virginia and the other southern colonies found the cultivation of tobacco their most profitable occupation. A considerable part of the people was engaged directly or indirectly in the tobacco business, either as planters, carriers, dealers, or exporters. So tobacco was generally acceptable; even the individual who neither smoked nor chewed accepted tobacco readily, since he knew he could always exchange it for other things.

The American Indians made, out of the inner whorls of certain periwinkle shells found along the seacoast, small beads, which they prized highly as ornaments. They were strung on thongs and braided and woven to make necklaces, bracelets, belts, and sashes, and the wealth and social standing of the brave was apt to be indicated by the quantity of this "wampum" which he could display. Being thus universally desired, wampum became the money of the Indians. So far we have merely another example of the ordinary historical origin of money. But note what happened when the white settlers came and opened commercial relations with the Indians. An important trade arose in skins and other forest products, particularly beaver skins, which the colonists could always export to Europe on favorable terms. Since an Indian would always accept wampum in exchange for beaver, the white traders gradually learned that they could safely accept it. And since the traders would always accept it, other whites who had dealings with them were willing to accept it. Before long any white man would accept wampum, knowing that he could always use it for making purchases, not only from the Indians but from the other whites as well. So there grew up a form of money, of no intrinsic value to the people themselves, but drawing its value entirely from its exchangeability, acceptable to each because acceptable to all, and based finally upon its acceptability by a race of barbarian neighbors.

The common characteristic. From these few examples one can derive a mental picture of the origin of money and of the great number of commodities which have served as money in various places and at various times. But what is of more significance,

one begins to sense a common feature among all this otherwise bewildering variety of monetary expedients, a feature which is at first likely not to be appreciated because of its very obviousness. This is the character of general acceptability in exchange for other things. Some of the early forms of money were clearly desirable for their own sake. Indeed all money doubtless had its origin in articles thus desirable. But others, such as the wampum used by the New England colonists, were of no use whatever for the direct satisfaction of their wants. Such forms of money were desired only because of their general acceptability in exchange.

It is the fact that money is acceptable generally which makes it acceptable to each individual, if we may be permitted a statement which may appear to involve reasoning in a circle. If the person who has valuable goods to sell is willing to accept in exchange a coin of gold, or a piece of paper engraved with the promissory note of a bank, or a bale of tobacco leaves, it is not ordinarily because he has any need of these particular articles; they may be quite incapable of satisfying directly any of his wants. He accepts them simply because he knows that at any time he chooses he will find other people willing to accept them in exchange for whatever they may have to sell to him. This is the fundamental characteristic of money; so long as this qualification is met there may be wide diversity as to other features.

Money is local. The general acceptability of money is always more or less local. Most American merchants would refuse to accept a Bank of England note in payment for goods sold, and a Canadian quarter is not generally acceptable in America except close to the border. The foreign traveler has to exchange his own money for the moneys of the particular countries which he proposes to visit. In general each nation has its own monetary system, and the money of one nation is usually not acceptable within the borders of other nations. By general acceptability we mean acceptability within the particular community.

Definition of money. We have now all the essentials of the scientific definition of money, which was stated but not explained in a previous chapter. *Money is anything which is generally accepted in a certain community in exchange for other things.*

Functions of money: Medium of exchange. The principal function of money is to facilitate exchange, in other words to serve as a *medium of exchange*. Without such a medium, exchange can operate only with great difficulty. The maker of shoes, for example, who wished to exchange some of his product for bread could succeed only when he had discovered the double coincidence of a baker who wanted to exchange bread for shoes. A similar problem would face the shoemaker in connection with every other good which he sought to obtain in exchange for his own product, shoes. To exchange so conducted the name of *barter* is applied. Barter undoubtedly prevailed under primitive conditions before the introduction of money, and it occasionally appears when the regular monetary system has temporarily broken down. But the modern economic order, based on division of labor and exchange of products, could never have become established had not the introduction of money released exchange from the restrictions and inconvenience of barter.

Standard of value. The second essential function of money is to furnish a standard of value. This function follows necessarily from the first. It is inevitable that the unit of whatever serves as the medium of exchange should be the unit for measuring the values of all exchangeable goods. Thus money furnishes the universal unit by which alone all kinds of wealth and services may be measured and compared and totaled. This function of money is as essential to the existence of the modern economic system as is the presence of a medium of exchange.

Store of value. Money performs also a third duty, in serving as a store of value. The significance of this function becomes apparent when we realize that money makes it possible for us to separate our sales from our purchases. The act of selling a pair of shoes, for instance, need not coincide with the act of buying a loaf of bread, as it would in a regime of barter. One can sell for money his services or what he produces and later according to his convenience buy the various things he needs. The business man can buy raw materials and pay for the services of the various factors of production before he sells the product of his industrial enterprise. In both of these cases money serves, not only as a medium of exchange,

but as a store of value. The consumer "holds" a certain portion of his wealth in the form of money, as does the producer. We shall see that the ability of money to serve in this fashion as a store of value is of great importance in connection with the determination of the rate of interest and the analysis of the trade cycle.

Government control of money: In general. While the historical data are not complete nor entirely unambiguous, it appears that money originated in each community naturally and unconsciously as the result of slowly established custom. That is, there was not at the start an inquiry and agreement upon the commodity which should serve as money. People simply fell into the habit of accepting some one commodity in preference to others in exchange for goods and services which they had to sell, much as boys at a certain stage in their development fall into the habit of using marbles as a medium of exchange. Thus the monetary system became firmly established in the mores of the community, and eventually the government took cognizance of it and proceeded to define and enforce it. For centuries the control of the monetary system has been jealously guarded by kings and rulers as one of the most precious attributes of sovereignty. Today the system in all its details is determined by law, and it is this modern legal monetary system with which we are concerned. Government (1) specifies exactly what kinds of wealth and property shall be money, fixes the monetary unit, defines the different kinds of money and their relations to each other, (2) coins the metallic money and engraves and prints the paper money, (3) in certain cases regulates the quantity of money, and (4) makes and enforces rules regarding legal tender.

Legal tender. While the general acceptability which is the essential characteristic of money grew up naturally through custom, modern governments have commonly found it desirable to enforce the acceptability of money by law. We thus have *legal tender*, which is defined as *any kind of money which according to law must be accepted when offered in payment of any obligation expressed in terms of the country's monetary unit*. When, as is usual, the monetary system consists of several kinds of money, the legal tender quality is often not the same with respect to the several kinds of money.

The legal tender distinction is purely a matter of legal enactment. It has nothing to do, as is sometimes supposed, with the question of intrinsic value. It is not concerned with contracts or other obligations not payable in money. Neither does it apply to contracts which by their terms are payable in a particular kind of money. There is nothing to prevent people making contracts payable in any kind of money they may agree upon. The normal purpose of legal tender laws is (1) to protect the creditor from being forced to accept anything but the best kind of money when different kinds are in circulation and (2) to protect the debtor against a capricious creditor who might refuse to accept a perfectly good kind of money or insist upon receiving some form which the debtor could not conveniently provide. It should be observed in passing that the government has the power, by altering the monetary unit or amending the law as to legal tender, to introduce unforeseen changes in the fulfillment of contracts, both public and private, previously made in terms of money.

Money and currency. Although, as we have noted, it is the principal function of money to serve as a medium of exchange, the fact is that many exchanges are consummated by means of instruments which are not in the strict sense money. Bank deposits — more strictly, individual deposits subject to check — are a medium of exchange, and indeed it appears that in the modern industrial nations bank deposits bear the larger part of the burden of this fundamental monetary function.¹ It has been estimated for example that in the United States some eighty or ninety per cent of the total volume of exchanges is conducted by means of checks or similar instruments which transfer bank deposits. Such instruments do not qualify as money under our definition because they are not “generally acceptable.” This makes it convenient to have another term, broader than money, which will include everything that serves as a medium of exchange. The term *currency* has been generally adopted. The complete picture of a *currency system* must include bank deposits and certain other forms of

¹ This statement is particularly applicable to the United States, Great Britain, and other Anglo-Saxon communities, where the deposit system of transfer by check is very highly developed. In the countries of Continental Europe there is relatively less reliance upon deposits and more upon money.

currency, such as postal money orders, express checks, etc. The subject of bank deposits will be treated at length in a later chapter.

In noting the great importance of bank deposits and other forms of currency which are not money, the reader will of course not forget that all forms of currency (except money itself) are themselves rights to receive money; in other words, that the whole currency system is based upon money. This point may be expressed in the statement that, although all currency performs the functions of a medium of exchange and a store of value, only money serves as a standard of value.

Varieties of money and monetary systems. Monetary experience has developed several kinds of money. The term *standard money* is used to designate those forms of money which are commodities or articles of wealth, having value, in their own right, apart from their use as money; in other words, standard money has "intrinsic value" equal to its face value. In modern times standard money has usually consisted of coins of gold or silver or of gold or silver bullion.

Standard money is usually accompanied in circulation by certain other kinds of money, whose value is not based upon the materials from which they are made but arises from the right of the holder to redeem such kinds of money in the standard money. To all kinds of money which are freely redeemable in standard money we give the title of *representative money*. The bulk of the representative money is usually in the form of paper notes or certificates, such as Bank of England notes or the notes of the United States federal reserve banks, though there is generally present also some representative money made of metal, such as the small change or token coins.

Money which is neither standard money nor representative money is called *irredeemable paper money*. Its value, not imparted by the material from which it is made, arises from the fact that it circulates and is generally acceptable in the community, reinforced sometimes by the hope of ultimate redemption in standard money. In practice such money is generally representative money in form, being government notes or bank notes whose promise to pay is not being honored.

A monetary system which consists of standard money or of both

standard and representative money is known as a *standard money system*. This is the system which generally prevailed in most of the civilized world for a century or two preceding the World War. In the United States a standard money system prevailed, with occasional interruptions, from 1792 until the abandonment of the gold standard in 1933.

Any money system not based on standard money may be classed as an *irredeemable paper money system*. Such a system contains neither standard nor representative money, and no form of money is redeemable in anything of intrinsic value. Historically irredeemable paper money systems have usually developed out of previously existing standard money systems. During the American Revolution an irredeemable paper money system prevailed, based upon the Continental bills of credit. The United States was again on an irredeemable money basis during the greenback regime, from 1862 to January 1, 1879. Many European nations went over to irredeemable money during the World War, and most of the nations of the world are, in one form or another, on that basis today.

Standard money systems: Principle kinds of money. The remainder of this chapter will be devoted to a more detailed consideration of the characteristics of standard money systems. In such a system there is usually in actual circulation a greater volume of representative than of standard money. Experience has developed several kinds of representative money, and the following classification of the component parts of a typical standard money system may prove useful:¹

1. Standard money
2. Representative money
 - (1) Standard money certificates
 - (2) Credit money
 - a. Government notes
 - b. Bank notes
 - c. Token money

A standard money certificate testifies to the fact that the government holds standard money which will be delivered to the bearer

¹ Classification must not be taken too seriously. It is never to be regarded as an end in itself; its purpose is merely to facilitate investigation. Classification there-

of the certificate on demand. It is thus of the nature of the warehouse receipt, and the reserve held by the government may never be less than the total amount of the certificates outstanding. The function is solely one of convenience; *i.e.*, to permit the use of the standard money without the necessity of handling the actual coins. The United States gold certificates, prior to 1933, were an example of standard money certificates.

The term *credit money* covers those forms which are of the nature of the promissory note. A *government note* is the government's promise to pay standard money or some other kind of money to the bearer on demand. A *bank note* is a similar promise of a bank. The United States notes (greenbacks) and the federal reserve notes respectively are examples in the United States.

Token money is credit money consisting of coins containing less than their nominal value of metal and redeemable in standard money. The token coins are usually made of silver or of other less precious metals, as copper, nickel, or bronze. When made of the same metal as the standard coins, the token coins are of proportionally less weight. Thus there is not as much silver in two half dollars, four quarters, or ten dimes, as in a silver dollar (the latter having formerly been a standard coin). Token coins are usually of smaller denominations than the monetary unit. Their function is to permit the "making of change"; *i.e.*, payments involving fractional parts of the monetary unit. Examples are the fractional coins of silver, nickel, and bronze in use in the United States. In the case of token money, the promise to pay is not stated upon the coin but is contained in the law requiring the public treasury to redeem the coin on demand of the bearer. Conversely the government will always give out the token coins on request in exchange for standard money.

Credit money (unlike standard money certificates) does not require the holding by the government of standard money equal in amount to the credit money outstanding; all that is required is a reserve sufficient to meet actual demands for redemption.

fore must always be more or less arbitrary, and it is not contended that the classification of money here presented is the only one that might be set up and defended. It does commend itself to us as on the whole the most useful for the present purpose.

The ideal money commodity. Standard money is a commodity. As we have seen, many different commodities have at various times and places been employed as money. Yet as time has gone on most of these monetary expedients have been gradually eliminated, till in modern times the civilized world quite generally settled down to use of the precious metals, gold and silver, as the standard money. The explanation of this important development will appear if we inquire what qualities a commodity should have in order to serve successfully as a standard money and note how well gold and silver meet the specifications.

First of all, the ideal standard money must have *general desirability*. The precious metals have always been desired for the fabrication of ornaments, and in modern times they have also important industrial uses, as in dentistry. In the second place, gold and silver have *great value in small bulk*. This is an important consideration, since a medium of exchange should be susceptible of convenient use and easy transportation. Consider for example the inconveniences which would accompany the use as money of such otherwise useful commodities as wheat, coal, or salt.

Thirdly gold and silver have the advantage of *durability*. Age and exposure do not materially affect them, and they may thus be kept indefinitely without danger of loss from physical deterioration. This is of course an important matter, if money is to perform its function as a "store of value." The Virginia colonists for example frequently suffered loss through the deterioration of their stores of tobacco money. Another desideratum of a convenient money commodity is *uniform quality*. All the possible advantages of monetary exchange are not obtained when each piece of money must be examined and tested as to its quality. Wheat would make a poor kind of money for this reason; so would coal or cotton or indeed almost any other possible commodity. The Virginia legislature had to take cognizance of the distinction between "good tobacco" and "bad tobacco" and provide officials to examine and grade the tobacco. Even so, the colonist had always to be on his guard against receiving payment in inferior or spoiled money. The precious metals, on the other hand, have perfect uniformity. When to this is added the quality of durability, which we have just noted,

it results that an ounce of fine gold is exactly like any other ounce of fine gold, whether it came from a California mine in 1849 or was mined last year in South Africa or the Klondike.

The convenience of a monetary commodity is greatly enhanced by *divisibility*. Thus gold and silver may be divided and made into coins of any desired size without loss of value. Gold and silver have also the obvious advantage of being easily recognized, of *cognizability*. They are not easily confused with other substances, and counterfeiting is difficult.

Types of standard money systems. There are many forms which a standard money system may take, but for our purposes it will be sufficient to distinguish four types for discussion in detail. They are the following:

1. The gold standard, or gold monometallism
2. The bimetallic standard, or bimetallism
3. The gold bullion standard
4. The gold exchange standard.

The gold standard. Under a gold standard the standard money consists of gold coins; all other kinds of money which may circulate as a part of the system (representative money) derive their value from the fact that they are redeemable in terms of gold coins. In such a system there is (1) a monetary unit, prescribed by law, (2) free coinage of gold, (3) legal tender of gold coins, (4) free circulation of gold coins, including the right to hoard, and (5) a free market in gold, treated as a commodity. The first two of these features require some further discussion.

The monetary unit. In a standard money system the law prescribes the monetary unit in which the various kinds of money are stated. When the United States was on the gold standard (from 1879 to 1933) the unit was the dollar, consisting of 25.8 grains of standard gold, standard gold being a mixture of nine tenths pure gold and one tenth an alloy of copper. The dollar thus contained 23.22 grains of pure gold and 2.58 grains of copper. Similarly the British monetary unit before September, 1931, was the sovereign, or pound, containing 123.273 grains of standard gold; British standard gold being eleven twelfths fine, the sovereign contained 113.001 grains of pure gold and 10.273 grains of alloy. Note that

the unit in a standard money system is always a certain quantity of metal whose weight and composition is exactly prescribed by law. The unit may or may not be represented by a coin. Very often the monetary unit is too small to make a convenient coin. This is of course the case in the United States.

Coinage. In the early days in California, after the great gold discoveries of 1849, gold became the medium of exchange by exactly the same natural process which led to the cattle money of the ancient pastoral tribes. Quantities of gold dust or nuggets were measured out in making payments. This was obviously inaccurate and inconvenient, and there was always the danger of misunderstanding and fraud. In the course of time business men and corporations made coins of various sizes, compositions, and designs, and the people found these coins much safer and more convenient than the gold dust money. These private coins were in their turn superseded by the gold and silver coins of the United States government. It is only when the government takes over the business of coinage and puts a stop to private coining that the ideal coinage system is obtained. Hence the logical reason for government monopoly of coinage, the development of which, it must be confessed, was furthered by the cupidity of many an early monarch.

Free coinage. The government obtains the metal for coinage in one of two ways; either it buys the metal in the open market or it allows *free coinage*. The former is simple enough, but the latter requires some explanation. Free coinage may be defined as *that system by which the government is legally required to coin for any person any amount of a particular standard metal which he may bring to the mint*. Free coinage is a necessary attribute of standard coin. For example there was in the United States until 1933 free coinage of gold.¹ Any person might bring to the mint any amount of gold of the standard prescribed by the coinage laws, and the mint would give back to him the same weight of gold coin. Governments sometimes perform this service for nothing; otherwise they make a charge sufficient to cover the cost of the coinage. In the United States the practice has varied. At first there was no charge ;

¹ Under the bimetallic standard or irredeemable paper standard to 1879 and the gold standard thereafter till 1933.

from 1853 to 1873 the charge was $\frac{1}{2}$ of one per cent; from 1873 to 1875 it was fixed at $\frac{1}{5}$ of one per cent; after 1875 there was no charge. When free coinage is granted without a charge for the cost of the operation, the term *gratuitous coinage* is employed. Note that there is free coinage, as defined above, whether the coinage is gratuitous or not.

The metal brought to the mint for coinage is often not of the standard composition; that is, not the exact mixture of pure gold and alloy prescribed by the law. In that case the mint or the government assay office will assay and refine it, charging the cost to the one who brought the gold or ore. The gold may then be returned to its owner or made into coin for him, and in the latter case there is a further charge for the alloy. In practice in the United States, the owner usually did not either take the gold or wait for its coinage, but took payment in lawful money; that is, sold the gold to the mint. The mint then held the gold either in "fine bars" for commercial use or in "mint bars" for coinage at its discretion.

The value relation between gold and money. Since the standard money under the gold standard consists of gold coins of an established weight and fineness, which the holder is free to hoard or melt down into bullion, the value of the standard money can never fall below that of an equivalent weight of gold in the form of bullion. On the other hand, if the value of gold coin were ever greater than the value of the corresponding weight of gold bullion, dealers, taking advantage of the right of free coinage, would buy gold in the commodity market and have it turned into coin until there was no longer profit in the operation because the values had become the same.

Under this system the various forms of representative money will have values equal to their nominal values because of the fact that they are all redeemable in terms of standard money, either by the government or by the agency which has issued them.

The bimetallic standard: General description. *Bimetallism*, though no longer in effect in any nation, has played an important rôle in monetary history and involves economic principles of great interest. In a bimetallic system both gold and silver are concur-

rently used as materials for the standard money, and the monetary unit is defined as either a certain weight of standard gold or a certain weight of standard silver, the law prescribing also the exact composition of each standard metal. The ratio between the weight of pure silver in the silver unit and the weight of pure gold in the gold unit is called the *mint ratio*, coinage ratio, or legal ratio. There is free coinage of both gold and silver, and the standard coins of either metal are unlimited legal tender.

In connection with bimetallism there arise certain important questions in addition to those which relate to monometallism. If debts may be paid either in gold coin or in silver coin and if either kind of coin may be freely obtained at the mint in exchange for the respective bullion, how will the people act? Will it make any difference which metal is used, and if so what considerations will determine the choice? Will both metals be actually used? These questions present an interesting theoretical problem; they have been the subject of violent political controversy in the United States, and they give special interest to research in monetary history.

The theoretical analysis essential to the answers to such questions is simple. It starts with consideration of the relation between the mint ratio and the market ratio, the latter being the ratio between the value of a given weight of gold and the value of the same weight of silver on the market. This ratio is of course the result of the respective values of gold and silver; these values are determined in their turn by the demand and supply in each case; and finally demand and supply depend respectively upon the uses of the metals for all purposes (not merely for money) and upon the conditions of mining. This ratio can obviously not be controlled by law, and it is subject to frequent change, as is the ratio between the values of any two commodities.

When the two ratios are equal. At any given time the market ratio must be either equal to, or less than, or greater than the mint ratio. Taking the first case, let us assume that a bimetallic system has a ratio of 16 to 1 and that the market ratio between silver and gold is also 16 to 1. A given amount of gold will exchange on the market for 16 times its weight of silver, and a given amount of

gold will make at the mint exactly the same value in coin as 16 times its weight of silver. There will be no reason for taking one metal to the mint rather than the other, and both metals will be coined.

When market ratio is less than mint ratio. Now let it be supposed that, the mint ratio still being 16 to 1, the market ratio is something less, say 15 to 1. Now anyone having silver bullion and desiring coin would find it to his advantage to exchange his silver for gold before going to the mint. For 15 ounces of silver he can buy 1 ounce of gold, which at the mint will give him just as many dollars as if he had taken there 16 ounces of silver. Therefore no one will take silver to the mint, and only gold will be coined. Furthermore there will similarly be advantage in melting down and using as bullion any silver coins which may then be in circulation. Silver coins weighing 15 ounces can be exchanged for an ounce of gold, which at the mint will be made into coin of the same face value as 16 ounces of silver coins; the exchange has netted 1 ounce of silver. Silver coins thus tend to disappear from circulation; gold is the only metal being coined, and the only coins in circulation are gold. The effect is practically the same as gold monometallism.

When the market ratio is the greater. The opposite results follow in case the market ratio is greater than the legal ratio. Suppose the legal ratio is 16 to 1 while the market ratio is 17 to 1. Anyone having an ounce of gold can now exchange it for 17 ounces of silver, 16 of which taken to the mint will yield him the same number of dollars as would have been obtained by having his ounce of gold coined. There is also gain to be obtained by melting down gold coin for bullion. Therefore only silver will be coined, and the gold coins already in circulation will tend to disappear. The effect is practically the same as silver monometallism.

Gresham's law. It will be noted that it is the metal which the mint overvalues in comparison with the market ratio which drives the other out of circulation. Coins of the first metal may be obtained from the mint more cheaply than coins of the other, for the reason that the mint places a higher value on that metal in terms of the other than does the market. This principle is part of a

broader generalization, known as "Gresham's law," which is fundamental to monetary theory and may be stated as follows: *When two or more kinds of money of unequal value are in concurrent circulation, each being available for payments, the inferior tends to drive the better out of circulation.* This law may be seen in operation whenever the conditions are fulfilled. It applies when anyone, selecting from a pocket book of bills, makes his payment with the worn and soiled bills and keeps those that are new and crisp, as well as when new heavy coins are driven out of circulation by those that are worn or clipped, and when the cheaper coins in the bimetallic system drive out the dearer. The better money may be hidden away and hoarded, or exported for payments abroad, or melted down for bullion. The inferior remains in circulation.

The reader scarcely needs the reminder that these results of a discrepancy between the two ratios will not come instantaneously. All persons concerned are not fully aware of these principles, the facts of the situation may not be recognized immediately, and there may be some who are not quick to take advantage of the opportunity thus presented. However since the actual business of taking metal to the mints and of using metal in the arts is in the hands of a few experts, this qualification is of little moment as regards the choice of the metal which shall be coined. The disappearance from circulation of the metal which is undervalued at the mint may take more time, though even this result will not be long delayed, particularly in a country like the United States where very little of the standard coin was ever actually used as pocket money. When most of the standard money is in the reserves of the banks and the government's vaults, it is certain that expert knowledge will quickly take full advantage of the situation.

Is a discrepancy between the ratios self-correcting? Before reaching our final conclusion we have also to observe that the forces set in operation by a discrepancy between the mint ratio and the market ratio have themselves a tendency to bring the ratios back to equality. Take the case of a market ratio less than the mint ratio between silver and gold. The existing silver coins begin to pass out of circulation, and new coins are made only of gold. The whole task of furnishing standard money, formerly shared by gold

and silver, is now about to be placed upon gold. This means an increased demand for gold and a consequent tendency for its market value to rise. On the other hand, silver being no longer used for standard money, the demand for silver is less, with a tendency to a decline in its value. The rise in the value of gold and the decline in the value of silver tend of course to increase the market ratio between them and so to correct the original discrepancy. It is possible theoretically that the equilibrium between the market ratio and the mint ratio might thus be restored before all the standard silver coins had been eliminated. From that point the coinage of silver would be resumed, and the bimetallic system would go on as before, except that some of the silver coins would be definitely lost and the circulating medium would contain relatively more of gold and less of silver than previously. The reader can readily enough work out for himself the opposite train of events that would be set in motion if the market ratio were greater than the mint ratio.

Exactly how great a discrepancy between the market and mint ratios could be overcome by the added demand for the overvalued metal depends among other things upon the relative conditions of production of gold and silver bullion and whether the burden is borne by a great nation or a small one or by few or many nations acting together. An international system of bimetallism including the majority of the leading nations, with agreement upon a mint ratio not far removed from the then existing market ratio, would undoubtedly have a considerable stabilizing effect on the market ratio.

The potency of this principle to perpetuate effective bimetallism in the face of fluctuation in the market ratio of gold and silver has however, in the heat of political controversy, been greatly exaggerated. At the time when the restoration by the United States of the double standard (abandoned in 1873) was being vigorously demanded, some of the extreme friends of bimetallism believed that, through the operation of this principle, bimetallism would always work and that neither metal would ever be entirely driven out of circulation, no matter what the relative conditions of gold and silver might be.

While it is difficult to derive a final answer to this question by means of pure theory, the world's experience with bimetallism furnishes an answer sufficient for all practical purposes. This and most of the other features of bimetallism are well illustrated by the experience of the United States.

Bimetallism in the United States. When the American people started their career as an independent nation, their money was a makeshift collection of paper notes and foreign coins left over from the colonial and Revolutionary periods. The greater part consisted of notes of the Continental Congress, of the several colonies or states, and of a few banks. Except for some silver shillings and sixpence previously coined by the Colony of Massachusetts Bay, there were no American coins. The most common coin was the Spanish peso or dollar. A national monetary system was first established by the mint act of 1792, only three years after the formation of the government of the United States. The unit was the dollar, of either gold or silver, though there was no one dollar gold coin. There was free coinage of both metals, and the coins, whether of gold or of silver, were unlimited legal tender. There was thus established a true bimetallic system.

Original coinage ratio. The coinage ratio was fifteen to one.¹ This ratio was chosen, following a careful investigation and report by Alexander Hamilton, with the intention of making it as nearly as possible the market ratio. As it turned out however the market ratio was actually about $15\frac{1}{2}$ to 1. In accordance with the principles with which the reader is now familiar, gold was not brought to the mint for coinage. Some silver was coined, but it happened that the new American dollars, though slightly lighter than the Spanish pesos, were found to be more attractive to the people of the West Indies and Central and South America because of their new and bright appearance. They therefore left the country almost as fast as they were coined, till their coinage was for this reason stopped in 1804. For a generation American coins made up only a

¹ The gold dollar contained 24.75 grains of pure gold and 2.25 grains of alloy, a total weight of 27 grains of the standard metal. The silver dollar contained 371.25 grains of pure silver and 44.75 grains of alloy; i.e., 416 grains standard. The coinage ratio was therefore $\frac{371.25}{24.75}$ or exactly $\frac{15}{1}$.

small part of the monetary system. Foreign coins continued to circulate, though the main reliance was always upon paper money; *i.e.*, government notes and bank notes.

The coinage ratio changed. This state of affairs was not considered satisfactory, particularly the absence of American gold coins. Recognizing the cause in the discrepancy between the mint ratio and the market ratio, Congress in 1834 reduced the weight of the gold coins and so changed the legal ratio to about sixteen to one.¹ The market ratio remaining close to $15\frac{1}{2}$ to 1, there was now a discrepancy on the opposite side. There was no longer inducement to bring silver to the mint; the previous tendency to export silver dollars was strengthened, and it was profitable to hoard or melt down such few silver dollars as were in circulation. Gold, on the other hand, was brought to the mint, and American gold coins began to take their place in circulation. In 1837 certain slight changes were made in the coins, the purpose being to make all coins exactly nine tenths fine. The mint ratio, though slightly reduced, still remained practically sixteen to one (to be exact, 15.988 to one), and this ratio has never since been changed.² Matters stood thus for a generation, bimetallism being legally in force but the practical effect being, from 1834 to 1862, gold monometallism. After the middle of the century, following the gold discoveries in California and Australia, great quantities of gold were coined.

The end of bimetallism. The greenback issues of the Civil War put the country on an irredeemable paper money standard from 1862 to 1878. In 1873 the standard silver dollar was dropped from the list of coins to be made at the mint. This action was taken as a

¹ The exact details of the change may be shown thus :

ACT OF	<i>Gold Dollar</i>			<i>Silver Dollar</i>			<i>Legal Ratio</i>
	PURE GOLD	ALLOY	STANDARD WEIGHT	PURE SILVER	ALLOY	STANDARD WEIGHT	
1792	24.75 gr.	2.25 gr.	27. gr.	371.25 gr.	44.75 gr.	416. gr.	15 : 1
1834	23.2 gr.	2.6 gr.	25.8 gr.	371.25 gr.	44.75 gr.	416. gr.	16.002 : 1

² The exact results of the act of 1837 are shown in the following table :

ACT OF	<i>Gold Dollar</i>			<i>Silver Dollar</i>			<i>Legal Ratio</i>
	PURE GOLD	ALLOY	STANDARD WEIGHT	PURE SILVER	ALLOY	STANDARD WEIGHT	
1834	23.2 gr.	2.6 gr.	25.8 gr.	371.25 gr.	44.75 gr.	416. gr.	16.002 : 1
1837	23.22 gr.	2.58 gr.	25.8 gr.	371.25 gr.	41.25 gr.	412.5 gr.	15.988 : 1

mere detail of a general revision of the coinage laws. It attracted little attention at the time and had no practical effect, since there was still no desire to bring silver to the mint for coinage into standard dollars. As a matter of fact neither gold nor silver was then in circulation. The act of 1873 thus definitely put an end to bimetallism and (after the resumption of specie payments, January 1, 1879) made the monetary system in law what it had long been in practical effect, gold monometallism.

Change in the market ratio. Conclusion. Now the act of 1873 had scarcely gone into effect when the market ratio between gold and silver, which for a century or more had remained almost constant (close to $15\frac{1}{2}$ to 1), began a spectacular rise. The very next year it reached the legal ratio of 16 to 1 and passed it; four years later it was 18 to 1, and in 1894 it reached 32 to 1, just double the legal ratio. Various causes coöperated to bring about this remarkable result, among them the decline in the use of silver currency in India and a number of the European countries, and the discovery of rich and extensive silver mines in America.

At this time the value of the greenbacks in terms of their face value in gold was gradually rising from about 90 in 1874 to par late in 1878. If the act of 1873 had not abolished the right of free coinage of silver, the time would have come (probably sometime in 1877) when silver dollars would have been cheaper than the greenbacks (by law redeemable in gold after January 1, 1879), and after the return to the gold standard silver dollars would have been far cheaper than gold coin. A flood of silver dollars would have inundated the currency, gold would have remained out of circulation, and practical silver monometallism would have prevailed. This causal relation was denied by certain of the advocates of bimetallism, who claimed, right up to 1896 and 1900, that the corrective principle to which we have already given attention would have brought the market ratio back to 16 to 1 before all the gold had been expelled. As to this the lesson of American experience is conclusive. From 1792 to 1873 the average annual market ratio between silver and gold was never below 15 to 1 nor above $16\frac{1}{4}$ to 1. During this time the mint ratio was either 15 to 1 or (practically) 16 to 1. The average annual discrepancy was

never as much as one point except in three years, and the maximum discrepancy was only $1\frac{1}{4}$, in the year 1813. Yet this small discrepancy was too much to be corrected by the free coinage of the cheaper metal, and throughout the whole period there was in practical effect either silver monometallism or gold monometallism, according as the legal ratio was below or above the market ratio. Certainly the restoration of the free coinage of silver could never have reunited the two ratios after their separation in 1874.

Bimetallism in Europe. Before the nineteenth century bimetallism had been the prevailing monetary system in Europe, silver being the commoner metal in circulation. England was the first to make a change; she definitely established her gold standard in 1816. In contrast the Continental countries generally continued their bimetallic systems through the first three quarters of the nineteenth century. We will find the French experience most instructive.

In France the mint ratio was $15\frac{1}{2}$ to 1. This was so close to the market ratio throughout the first three quarters of the nineteenth century that the French bimetallic currency never entirely lost either metal until the free coinage of silver was legally terminated in 1873. During the greater part of the period from 1820 to 1850 the mint ratio overvalued silver, which tended as a result to flow into France. Gold was flowing out. The net result was a large increase in the stock of specie of both kinds taken together, necessitated by the concurrent growth in France's population and wealth. Before gold had quite vanished from circulation however the situation abruptly reversed itself. After 1850 unprecedented supplies of new gold were poured into the world's markets from California and Australia. The market ratio fell below $15\frac{1}{2}$ to 1. A great influx of gold took place, and a corresponding, though by no means equal, outflow of silver; France was still increasing her total metallic stock. But the production of gold reached its maximum about 1860. Thereafter it barely held its own, while a few years later there began the striking increase in the production of silver which we have already noted. The market price of silver fell slightly about 1865. Silver ceased flowing out of France and began to flow in. For a few years the market ratio was almost exactly equal to the mint ratio. Then in 1873 began the revolutionary decline in the

market price of silver, which quickly brought the market ratio to 16 to 1. Gold began to flow out in large quantities, while silver flowed in. Not wishing to lose her gold — gold by this time having gained general acceptance as the preferable monetary standard — France in that year stopped the free coinage of silver.

Triumph of the gold standard. The newly created German Empire adopted the gold standard in 1873. The Scandinavian countries followed suit in 1875. The United States, France, and other bimetallist countries — notably the other members of the Latin Monetary Union — did the same in effect (although formal adoption came later) by suspending the free coinage of silver. England, as we have seen, had been on the gold standard since 1816. When the majority of the leading commercial nations had adopted the gold standard, the others were hastened into line by the further advantage of uniformity, and the swing to the gold standard acquired world-wide momentum in the last quarter of the nineteenth century. By 1900 all important countries in Europe and many others had currencies either based on gold or related to gold.

The “limping standard”: In general. In spite of the widespread adoption of the gold standard the problem of silver money was not laid to rest. On the contrary, it has recurrently demanded attention ever since silver was demonetized, particularly in the United States. When in the period 1871 to 1874, the leading nations of the world definitely abandoned bimetallism and embraced the single gold standard, the former standard silver coins were left in a peculiar position. Their free coinage privilege was gone, and they were therefore no longer true standard money. At the same time they were generally left unlimited legal tender. Within a few years the value of silver had so fallen with respect to gold (*i.e.*, the market ratio had so risen) that the silver in these coins was no longer worth as much as the face value of the coins. The coins no longer had full intrinsic value, as is the case with real standard money. These hybrid coins — neither true standard money nor yet real token coins since they were not redeemable in standard money — continued thus to circulate side by side with the standard gold coins and gave rise to the contemptuous title, the “limping standard.”

American silver dollars after 1873. Whereas in Europe, as we have seen, silver coins were generally more numerous than gold at the time when bimetallism was given up, in America it will be recalled that this was not the case. For more than a generation (except for the greenback interregnum) the United States had had virtual gold monometallism, and even before 1834 the coinage of silver dollars had been small, and many of the coins made had been exported. In the whole history of the country up to 1873 the coinage of silver dollars had amounted to only \$8,000,000. Silver dollars were thus in 1873 an insignificant part of the circulating medium, and the problem of the limping standard need never have been acute in America, had it not been for subsequent legislation of a highly artificial sort.

Reference has been made to the strong political demand for the restoration of the free coinage of silver at the legal ratio of 16 to 1 which developed in conjunction with the remarkable fall in the price of silver beginning in 1873 and continuing to the end of the nineteenth century. This movement did not succeed, but it did cause Congress to enact two important measures of a compromise nature.

Bland-Allison and Sherman acts. In the year 1878 the Bland-Allison act compelled the Treasury to purchase each month \$2,000,000 worth of silver bullion and coin it into "standard" silver dollars. One year's coinage under this law was three times the total amount of silver dollars coined in the entire history of the country up to that time, and the law remained in effect for twelve years, leading to the coinage of 378 million silver dollars. Then in 1890 the Bland-Allison act was superseded by the Sherman act, which increased the Treasury's monthly purchases of silver to 4,500,000 ounces. This silver was paid for by the issue of legal tender treasury notes redeemable in either gold or silver coin at the option of the Treasury. The silver so bought was held by the Treasury, being coined into dollars only as required to redeem the treasury notes.

The Sherman act, though differing in detail from the Bland-Allison act, had nevertheless the same fundamental effect. Both acts inflated the monetary system with silver dollars, or notes backed primarily by silver dollars, whose metallic content was

worth less than — by 1894 only half — the standard gold dollar. Such a movement, continued long enough, would have had the same effect as the free coinage of silver; *i.e.*, the expulsion of gold and the substitution of the silver standard. This effect was already being felt. Gold was rapidly disappearing from circulation, accompanied by heavy exports to foreign countries. The silver dollars and the treasury notes of 1890, while never formally at a discount, were more and more regarded with suspicion, and the Treasury was hard put to it to maintain the stability of the monetary system and meet its gold obligations. The monetary inflation and weakness constituted one of the chief causes of the disastrous panic of 1893. Inflation was finally stopped by the repeal of the Sherman act in 1893.

The free silver party was defeated in the presidential elections of 1896 and 1900, and in the latter year Congress definitely established the gold standard. In the same year Congress directed the conversion of the treasury notes of 1890 then outstanding into silver certificates backed by coinage of an equal quantity of silver dollars; the total coinage under the Sherman act, when thus completed in 1905, amounted to 187 million dollars. The heritage of the nineteenth century was a total coinage of 578 million silver dollars, almost all under the artificial stimulus of legislation subsequent to the demonetization of silver.

The Pittman act. At the time of the World War the United States lost an excellent opportunity to get rid of the “limping” silver dollars without loss. The sudden extraordinary rise of all prices finally brought the price of silver to the point where the market ratio was less than 16 to 1. For several months in the winter of 1919–1920 there was thus more than a dollar’s worth of metal in a silver dollar, and there was profit in melting down the silver coins. Some were melted and exported. Moreover there came a special demand for silver from the British and Indian governments. Congress, through the Pittman act of 1918, permitted the Treasury to melt and sell the silver dollars held back of the silver certificates up to 350 millions, the corresponding silver certificates being replaced by a special issue of federal reserve bank notes. A little over 270 million dollars were so melted down. But the Pittman act

provided that as soon as the emergency was over the silver dollars must be replaced by new dollars coined from American silver purchased by the government. Recoinage of the 270 million silver dollars was commenced in 1921 and completed in April, 1928.

The gold bullion standard. A modification of the gold standard was employed for a time after the World War by France and England. For example England, having departed from the gold standard during the War, returned in 1925, not to her former standard, but to what is known as the *gold bullion standard*. The right of free coinage of gold was given up (except in the case of gold bullion brought to the mint by the Bank of England), and the Bank of England notes and other representative money were denied the right of redemption in gold coin. Instead there was substituted redemption by the Bank of England of its notes in gold bullion at the price of £3 17s. 10½d. per standard ounce, but only in large bars containing approximately 400 ounces of fine gold. The Bank was also required to purchase gold bullion whenever offered at the price of £3 17s. 9d. per standard ounce.

We have here illustrated the essentials of the gold bullion standard. There is not free coinage and little or no circulation of standard coins. Representative money is redeemable, not in gold coin, but in gold bullion. Redemption is usually limited to such large quantities of gold that the right is in practice available only for gold export or for very large domestic transactions. The fact remains however that there is redemption in the standard metal, and this fact is sufficient to keep all forms of representative money at par with gold. Free trading in gold bullion is permitted, but the absence of standard coin and the limitation of redemption to such large quantities of money virtually prevents hoarding.

The gold exchange standard. During the period following the World War, another variation of the gold standard was extensively adopted in Europe, especially by small countries. In order to avoid the necessity of holding a large supply of gold as a basis for the currency, these nations made their representative money redeemable, not in gold, but in exchange upon some country having the gold standard. When an individual presented a sum of representative money for redemption he would receive foreign exchange, which

could be used to obtain gold from abroad. The government itself treated bank deposits in such financial centers as New York, Paris, and London as parts of its own monetary reserve or standard money. This is the *gold exchange standard*.

It will be noted that under this standard the currency of a foreign country becomes the standard money. The government is obligated to redeem representative money either in gold or in foreign exchange at its own option. It is usual for the government to keep part of its reserves in the form of gold and part in the form of foreign exchange. There is of course no free coinage of gold, and little or no circulation of gold coin, but redeemability in gold exchange keeps the various forms of representative money at parity with gold. When a number of countries employ the gold exchange standard, the gold reserve of one country may serve at the same time as the gold reserve for several others. Should all of the countries on the gold exchange standard experience a drain of gold at the same time, their withdrawals of foreign exchange from the financial centers would impose a grave strain. Therefore the wide employment of the gold exchange standard introduces an element of precariousness into the international monetary system.

EXERCISES

1. The silver dollar contains 371.25 grains of pure silver. There are 480 grains in an ounce.

- (a) The market price of silver today is about 43 cents per ounce. What is the intrinsic value of the silver dollar?
- (b) At what price of silver would the silver in a dollar actually be worth a dollar?
- (c) Suppose silver dollars were recoined into an equal number of dollars of lighter weight, and the surplus silver sold as bullion. How would this affect their intrinsic value? their value as coins? How would it affect silver producers? industrial consumers of silver? Explain, in each case.

2. The figures below indicate the weight of pure gold in the gold dollar at different periods in our monetary history :

1792-1834	24.75 grains of pure gold
1837-1933	23.22 grains of pure gold
1934-	13.71 grains of pure gold

- (a) Determine the mint price per ounce for gold, during each of these periods.

-
- (b) Daily quotations are to be found in the newspapers for silver, but no quotations for gold. Why?
 - (c) Prior to 1933, Congress had fixed the weight of the gold dollar at 23.22 grains of pure gold, and the weight of the silver dollar at 371.25 grains of pure silver. The price of gold was fixed, while the price of silver fluctuated. How was this possible?

3. After 1873 the market ratio rose as high as 80 to 1. At that ratio what was the market price of silver? Why did not silver coin drive out gold coin and leave the United States on a silver standard?

4. There are 480 grains in an ounce. What would be the weight of the gold dollar if the mint price for gold were fixed at \$20 an ounce? \$24 an ounce? \$30 an ounce? \$35 an ounce?

XXIII

IRREDEEMABLE MONEY SYSTEMS

Origin and general features. The distinguishing characteristic of an irredeemable paper money system is the absence of any standard money whatever. And by definition there can of course be no representative money. The monetary authority, whether it be the government or the central bank, no longer redeems money either in a commodity of intrinsic value or in the money of any other nation.

While it is perhaps conceivable that an irredeemable paper money system might be originally established without reference to a previously existing standard money system, this has as a matter of history never occurred. Actually the origin is to be found in the suspension of specie redemption of the representative elements (generally government notes or bank notes) in a standard money system. The operation of Gresham's law forces out of circulation the standard money and any other forms of money which may still be redeemable, and the irredeemable paper money thereafter becomes the basis of the system.

Up to the time of the great depression (beginning in 1929), nations had rarely given up their standard money systems except under the pressure of war or of some other crisis. Departure from the standard money system was a symptom of financial weakness. In recent years however many nations have abandoned the gold standard in favor of irredeemable money systems, not because they were actually forced to do so or because their metallic reserves were in danger of exhaustion, but for reasons of deliberate policy. In subsequent sections we shall give attention to the motives which may lead a nation to abandon the tie with gold.

The irredeemable money system, like the standard money system, may comprise several different kinds of money. The basic money will generally be either government notes or bank notes. Frequently both of these forms of credit money will be in use, the

bank notes being redeemable in government notes. Token money is generally present, performing the same function as in a standard money system, of course not redeemable in standard money, though it may be redeemable in other forms of money of larger denomination. The token money may be paper notes instead of coin. Bank deposits of course play the same rôle as in a standard money system, deposits being payable in the irredeemable paper money.

It may at first sight be hard to understand how an irredeemable money system can have any value at all, or why anyone should consent to receive such money. Yet that this does occur is demonstrated over and over in the annals of the world's monetary history. Irredeemable money, as we have seen, has almost always been representative money which has ceased to be redeemable. The people having become used to the credit money, force of habit continues its use even when redemption has ceased and the chance of future redemption has become virtually nil. Even when it declines in value, as evidenced by rising prices, it continues to circulate. Each individual feels that he can accept the money so long as he can pass it on to others, and so the one essential characteristic of money, general acceptability, is maintained. It is remarkable how far habit will thus lead a people. The Continental bills of credit of the American Revolution circulated till they were worth less than one cent on the dollar. The common people of Germany after the World War were still using their paper mark when its ratio of depreciation was impossible to calculate and when it was finally stabilized at one trillion to one!

American experience to 1879: Before the Civil War. The general nature of irredeemable paper money systems is well illustrated by American experience. Long before the Revolution the American colonies had saturated their monetary system with notes of the several colonies and of certain so-called banks. These notes were originally redeemable in coin, or intended to be so. But in most of the colonies the practice of redemption did not long continue, and with suspension of specie payment there were generally large increases of the amounts outstanding. By the time of the Revolution irredeemable paper money systems were quite generally established, although there was some circulation of foreign coins

and Massachusetts Bay silver coins and the use of certain other commodity moneys, of which some examples have been noted.

The War of the Revolution was financed principally by the issue of the promissory notes of the Continental Congress, which soon flooded the circulation, driving out coin and becoming the regular money of the people. These notes purported to be credit money, the promise of the Continental Congress to pay coin. But there was no coin for that purpose, the notes were from the first not redeemable on demand, and no one knew when they would be or if they would ever be redeemed. They were issued in large quantities; their depreciation was prompt and disastrous. Eventually, having declined to less than one cent on the dollar, the people lost all faith in them and refused to accept them. Gold and silver coins then gradually found their way back into circulation. In the financial reconstruction after the adoption of the Constitution the notes were redeemed, so far as presented, at the rate of one cent on the dollar. Thus ignominiously ended the first national experience with irredeemable money.

The closing of the First United States Bank in 1811, followed by the War of 1812 and the panic of 1814, led to suspension of specie payments by the United States and by the local banks, which sprang up in great numbers to fill the vacuum left when the government bank passed out of existence. The country was on an irredeemable money basis, consisting chiefly of treasury notes of the United States and local bank notes, until the establishment of the Second United States Bank in 1816. There were the usual accompaniments of depreciation, price disturbances, and speculation.

The Second United States Bank went out of existence in 1836, there was a panic in 1837, and for a short period irredeemable money again prevailed.

The greenback régime. The Civil War was financed in large part by the issue of government notes — the United States notes or greenbacks. The greenbacks, though promissory notes of the United States, were not at the outset redeemable in standard money. It was generally supposed that the notes would be redeemed in gold coin as soon as the war was over, but this was not

finally done until January 1, 1879. From the start the greenbacks depreciated and drove most of the gold and silver coin out of circulation. The other kinds of money in circulation were either other classes of notes of the United States or the notes of banks, themselves redeemable in greenbacks. The greenbacks were thus the basis of an irredeemable paper money system for the seventeen years from 1862 to 1879, though always somewhat supported by the hope of ultimate redemption. So low did they fall (the lowest point was about thirty-five cents on the dollar) that the small silver coins, though reduced in weight seven per cent in 1853, became more valuable as bullion than as coin and generally disappeared from circulation. Congress found it necessary to provide paper token money (notes) which should not be more valuable than the greenbacks themselves; these soon acquired the euphonious title of "shin plasters." During this period the usual economic and social consequences of a depreciated irredeemable money system were experienced, especially during the earlier years.

After the restoration of gold redemption on January 1, 1879, the greenbacks became part of the standard money system of the United States. In 1878 Congress forbade further retirement of these notes and required that as they were redeemed they should be paid out and so kept in circulation. This fixed the amount at \$346,681,-016, and the greenbacks in this amount have remained a part of the nation's monetary system to the present day.

Irredeemable paper money since 1914: General desertion of the gold standard. Lessons similar to those taught by the Continental bills of credit and the greenbacks in America may be drawn from early paper money experiences in a number of other nations. Our present investigation however leads us now to an examination of the period from 1914 to the present, a period during which revolutionary changes occurred in the monetary systems of most of the leading nations of the world.

It will be recalled that prior to 1914 the gold standard had been almost universally adopted throughout the world. Upon the outbreak of the World War all European countries, neutrals as well as belligerents, abandoned the gold standard in order to safeguard their metallic reserves, and most other countries followed suit.

Irredeemable paper money systems, consisting generally of the notes of central banks, prevailed throughout most of the world.

The period of recovery between 1923 and 1929 witnessed a concerted return to the gold standard, which, in a world forcibly educated in the evils of wartime inflation, was generally believed to promise stability and so to represent the desirable norm. Certain small countries — for example, Turkey and Spain — failed to take part in this restoration movement, but some thirty other countries did so. England returned to gold (gold bullion standard) in 1925 at the pre-war valuation of the pound. Many other nations, such as France, Italy, and Belgium, reduced the gold content of their monetary units upon returning to the gold standard.

The depression overthrew the gold standard again. In 1929 and the years following, one country after another — more than thirty countries by March, 1934 — abandoned gold, usually for the second time. England stood out for many months, but was forced off gold in September, 1931. The British dominions, the Scandinavian countries, and certain others subsequently regulated their currencies so as to maintain a determined relationship to the pound sterling and thus came to be known collectively as the sterling area. Only a small group of nations — France, Belgium, Holland, Switzerland, Italy, and Poland — remained to constitute the gold bloc,¹ resolved to defend the gold standard against the forces threatening its complete overthrow.

The present British monetary system. As an example of present European monetary systems, we may note that England finally abandoned her gold bullion standard in September, 1931, by repeal of the law requiring the Bank of England to redeem its notes in gold bullion at the fixed price of £3 17s. 10½d. The paper money thus ceased to be redeemable in gold. No embargo was placed on gold exports, and a free market was left for gold. Gold immediately went to a premium, and it has been so ever since. Its market price on October 19, 1938, was £7 5s. 8½d. Of course it fluctuates from day to day and during the day. Obviously no one has any interest in exercising his legal right to sell gold to the Bank in exchange for

¹ Germany and a number of central and eastern European nations remained nominally on gold but imposed restrictions on foreign exchange.

notes at the price of £3 17s. 9d.¹ It is evident that England has now an irredeemable paper money system.

Recent American monetary experience: 1914–1933. The United States was one of the few nations that adhered to the gold standard during the World War, though an embargo on gold exports was enforced from late in 1917 to the summer of 1919, and while redemption of other forms of money in gold was not refused, it was discouraged. Not only did the United States thus withstand the forces working against the gold standard, but during and after the war her financial strength in comparison with the rest of the world was the occasion of an unprecedented flow of gold into this country. The gold stock of the United States increased in ten years from a little less than two billion dollars to more than four and a half billions, or nearly half the total gold stock of the world. During the first two years of the depression there was further increase, to a total of almost five billions. The depression did not develop weakness in the American monetary system as such. Trouble in this country arose out of defects in the banking system — structural and functional defects of long standing, which came to a head, together with the various financial disturbances here and abroad, in connection with the panic of 1929 and the following years of depression. This record must be postponed until our study of banking in later chapters. In the present connection it is to be noted that failing confidence in the American banks had progressed to a state of acute crisis late in 1932. Runs upon the banks, withdrawal of money for hoarding by frightened depositors, and wholesale bank failures led the governors of several states to order all their banks closed in February and March of 1933. On March 6, only two days after his inauguration, President Roosevelt by executive proclamation closed all the banks in the United States.

Bank holiday and gold embargo of March, 1933. The proclamation of March 6, 1933, went further and forbade, except with special government license, the export of gold or silver or the paying of gold or silver by any bank. Simultaneously the United States Treasury ceased redemption of gold certificates and all gold payments. Three days later the emergency banking act of March 9 confirmed

¹ See Chapter XXII, page 519.

the President's proclamation and further authorized the President to regulate and prohibit hoarding by individuals of gold or silver and to require all persons to surrender to the United States Treasury all gold coin and bullion and gold certificates in exchange for other kinds of money. These provisions as to gold were immediately put into effect. It was generally expected that at the end of the banking crisis the nation would return promptly to the gold standard. Acceptance of this understanding is demonstrated by the fact that foreign exchange rates registered no depreciation of the dollar for more than a month.

Official abandonment of the gold standard. Actually, while the acute banking crisis was virtually over in a week, no move was made to restore the gold standard, and on April 19, 1933, the President proclaimed an embargo of indefinite duration on gold exports, having previously extended the regulations against hoarding of gold by individuals. This was formal notice of indefinite abandonment of the gold standard. The proclamation was immediately reflected by depreciation of the dollar on the foreign exchange markets of the world. For example France was on the gold standard, and the franc had a par value of 3.92 cents in American gold money. The exchanges had registered no substantial departure from this valuation between March 6 and April 19. On the latter date, the exchange rate began rising, and by October, 1933, it had reached 5.8 cents, indicating a depreciation of the dollar of almost a third.

The concluding act came on June 5, 1933, when the President approved a joint resolution of Congress which canceled the gold clause in all federal and private obligations, making them as well as all other contracts and debts payable in any form of legal tender. This action led to lawsuits ultimately involving some 100 billion dollars in gold-clause obligations. On February 18, 1935, the Supreme Court rendered a 5 to 4 verdict in favor of the government, holding that creditors must accept in depreciated currency, dollar for dollar, the sums named in their contracts, and that, while the government's action in repealing the gold clause in its own contracts was unconstitutional, plaintiff had no case against the government because no actual damage had been

shown. The government subsequently enacted legislation designed to guard itself against further suits in this connection.

Devaluation of the dollar. The monetary policy of the Administration had two announced objectives: (1) to secure greater freedom to pursue an "easy money" policy, and (2) to raise the internal price level, especially the prices of staple commodities dealt with in international trade, such as wheat and cotton; the price level prevailing in 1926 was evidently regarded as the goal. To accomplish these ends it was considered necessary first to cut the monetary system loose from gold and then to depreciate the value of the dollar.

Depreciation of the dollar was sought by means of an extensive program of gold purchases, begun late in October, 1933. The offer by the government to purchase gold with dollars was of course equivalent to an offer to sell dollars for gold. It is not necessary to enter here into the details of this maneuver. Suffice it to say that the government began to buy newly mined domestic gold at \$31.36 an ounce, far above the old gold-standard price of \$20.67 an ounce and 27 cents above the then current price in the world market. Shortly thereafter this program was extended to include buying gold abroad, and the buying price was successively raised until it reached \$34.45 an ounce on January 15, 1934. The number of dollars actually sold by the government for gold during these months was only about 130,000,000, but the action of speculators swelled the total sales of the dollar during the period to an amount estimated at nearly a billion. Although the depreciation of the dollar exerted some effect upon commodity prices, the result fell far short of what was intended.

Gold reserve act and proclamation of 1934. The gold-purchase program was followed on January 30, 1934, by the far-reaching gold reserve act and the accompanying proclamation of January 31, which devalued the dollar and nationalized gold. These measures provide that there shall be no gold coinage, no circulation within the United States of gold or gold certificates, no free redemption of currency in gold, no dealing in or holding of gold except under Treasury license; they empower the Treasury to deal in gold so as to maintain the parity of all kinds of currency and provide that

gold bullion shall be made available under Treasury license for the settlement of international balances and for domestic use commercially and in the arts. The Treasury buys all gold presented, except such as may be known to have been hoarded in violation of the orders established in 1933, subject to the usual mint and handling charges. Hoarded gold remains subject to confiscation with a 100 per cent penalty charge, although in practice the Treasury has been purchasing such gold at the former mint price of \$20.67; up to the end of July, 1935, the Treasury had received 31 million dollars of such gold and 97 million dollars of the old gold certificates. The security for gold certificates, which now call for payment to bearer on demand "as authorized by law," and the reserve for United States notes and the treasury notes of 1890 are held in the form of bullion, since gold coinage has been stopped and all gold has been converted into bars. The new gold certificates are simply held by the federal reserve banks as final reserves, redeemable at the Treasury as needed for the settlement of foreign balances. Title to all gold coin and bullion held by the federal reserve system is transferred to the Treasury, in return for an equivalent credit in terms of the old gold content of the dollar payable in gold certificates. Federal reserve notes are made redeemable in lawful money, and gold certificates are substituted for gold in all the reserve and collateral requirements of the federal reserve system.

In addition to a grant of further powers relating to the monetary use of silver, the remaining provisions of the gold reserve act were concerned with the fixing of the gold parity of the dollar, giving a semi-definitive stabilization of the dollar in gold at a depreciated or devalued level. This act amended the 50 per cent devaluation authority given the President in the "Thomas amendment" to the emergency farm relief act of May 12, 1933, by providing that any weight of gold fixed for the dollar should not be more than 60 per cent of the then existing weight and granted the President continuing power to make successive revaluations within the specified limits for a period up to three years. It was directed that any increase in the value of gold held by the Treasury resulting from a reduction in the weight of the gold dollar should be carried in the Treasury as a miscellaneous receipt and that, in the event of any

increase in the weight of the gold dollar, the reserve for notes and the security for gold certificates should be correspondingly augmented by a transfer of gold from the general fund of the Treasury.

New monetary unit. Legal tender. The President immediately fixed the gold content of the dollar at $15\frac{5}{16}$ grains, $\frac{9}{16}$ fine, the equivalent of an even \$35.00 an ounce as the dollar price of gold.

As the reader will recall, the gold dollar, since 1837, had remained fixed at a pure gold weight of 23.22 grains and a standard weight (*i.e.*, $\frac{9}{16}$ gold, $\frac{1}{16}$ alloy) of 25.8 grains. As now fixed, the dollar contains $13\frac{5}{8}$ (approximately 13.7) grains of pure gold and has a standard weight of $15\frac{5}{16}$ (about 15.24) grains. This is a reduction of the dollar to 59.06 per cent of its former weight.

The act of June 5, 1933, which abrogated the gold clause in public and private contracts, enacted that "all coins and currencies of the United States (including federal reserve notes and circulating notes of federal reserve banks and national banking associations) heretofore or hereafter coined or issued, shall be legal tender for all debts, public and private. . . ." This makes all forms of money, apparently even the small-change coins, unlimited legal tender.

"Gold profit" and stabilization fund. The dollar value of the Treasury's gold was by this action increased by an amount slightly in excess of two and three-quarters billion dollars. The federal reserve banks were left with a claim to as many dollars' worth of gold as before devaluation. But since the new dollar contained only 59 per cent of the former weight of gold, the banks' share of the total gold in the country's monetary stock (measured either by weight or by dollar's worth) was reduced to 59 per cent of what it had been.

The gold reserve act gave the Secretary of the Treasury power for a period up to three years to deal in gold and foreign exchange and any other necessary instruments for stabilizing the exchange value of the dollar. A fund of two billion dollars was provided for this purpose, to be taken from the increased value of the Treasury's gold which might result from devaluation, with the permission that any portion of the fund not being currently used in stabilization operations might be applied by the Secretary of the Treasury to the

purchase of government obligations. The two billion dollar exchange stabilization fund was not set up out of the Treasury's gold profit until the end of April, 1934, when 1,800 million dollars were transferred to that title on the books of the Treasury from "Gold in General Fund" and 200 million dollars were placed to the credit of the Treasury in the Federal Reserve Bank of New York.

Silver policy: The facts. With the abandonment of the gold standard, the silver issue, which we have seen playing a leading part in American politics in the last quarter of the nineteenth century, once more achieved prominence. Agitation for increased recognition of silver in the monetary system was insistent and led to a number of legislative enactments. The Thomas amendment to the emergency farm relief act in 1933 authorized the President to accept silver up to two hundred million dollars' worth on the basis of fifty cents per ounce in payment of the allied war debts due in 1933 and to issue silver certificates against the silver so received. A number of payments received in 1933 were thus made in depreciated silver. In the ill-starred world monetary and economic conference in London in July, 1933, an agreement was negotiated designed to raise and stabilize the price of silver. As its share of the plan the United States agreed to purchase twenty-four and a half million ounces of silver a year. In effect the agreement called upon this country to protect the price of silver against whatever damage might be done it by the limited disposal of some of India's surplus stock. In December, 1933, the President, acting under the provisions of the gold reserve act, announced that the United States mints were open, during the next four years, for the purchase of all the silver mined in the country, at the statutory price (unchanged since 1837) of approximately \$1.29 per ounce, the producer however to receive only half of this amount, or approximately 64.5 cents per ounce, while the balance represented a 50 per cent *seigniorage charge* or profit to the government.¹ This was 21 cents above the current market price at that time. The official price of newly mined silver remained at 64.5 cents until April, 1935, when, to keep the price offered to domestic producers above the rapidly rising world price,

¹ A charge made by a government for coining bullion is called *seigniorage* if it exceeds, and *brassage* if it does not exceed, the actual expense of coinage.

successive proclamations two weeks apart raised it first to 71.11 cents and then to 77.58 cents an ounce.

Demand for further action on behalf of silver persisted and came to a head in June, 1934, when Congress passed the silver purchase act. The Secretary of the Treasury is directed to buy silver as he may deem it advisable at a price not above its monetary value (\$1.29), except that in the case of silver in the United States prior to May, 1934, the price may not exceed \$0.50. The Secretary may sell silver whenever its market price is above its monetary value. The purchases are to continue until silver constitutes 25 per cent of the monetary stock of the country. Silver certificates, now made full legal tender, must be issued at a face value not less than the cost of all the silver bought, and they may be issued in an amount not larger than the statutory valuation of the silver.

By the same act the President was given power to combat the hoarding of silver by nationalizing the commodity, which he did by executive order in August, 1934. The order required the delivery of silver bullion to the Treasury within 90 days at \$0.50 an ounce. The Treasury promised to see that adequate supplies of silver would always be available at a fair price for legitimate commercial purposes.

Monetary significance of silver. So far as the American monetary system is concerned, the action taken with regard to silver has thus far been of little significance. The United States Treasury has acquired a considerable amount of silver bullion; it held 895 million ounces on January 31, 1939. There has been a moderate increase in the circulating medium through issue of silver certificates. In June, 1934, the Treasury began their issue, backed by silver at the statutory valuation of \$1.29 an ounce, against 10 million coined silver dollars then held free and against 62 million ounces of silver which the Treasury had acquired at a cost of 47 million dollars from foreign debt payments and the purchase of newly mined domestic supplies. These 62 million ounces, at the statutory valuation, permitted an issue of about 80 million dollars of silver certificates. Purchases by the Treasury were curtailed as reduction of available stocks and resulting speculative activity produced a 45 per cent increase in the price of bar silver in the first four months of 1935.

On January 31, 1939, there were 1,568 million dollars of silver certificates outstanding, an increase of 1,086 millions since February 28, 1933. The stock of silver dollars (and bullion), above the requirements for backing the certificates increased only 57 million between these two dates. The total effect is an increase of 1,143 millions in the country's monetary stock on account of silver. During the same period the increase in the total monetary stock was 11,902 millions. Silver accounted for one-tenth of the increase. Of the silver bullion held by the Treasury on January 31, 1939, there were only 136 million dollars' worth not already pledged to redeem outstanding silver certificates.

The announced one-to-three ratio between silver and gold in the monetary stock is of no immediate practical significance. The present stock of silver dollars and bullion is far short of the amount required and is not likely to reach the announced ratio in the near future. And in any event the ratio means little in a monetary system which has largely lost connection with the Treasury's stock of gold. Should there be a return to the gold standard, an increase in the ratio of silver to gold would serve merely to increase the "limp" of the "limping standard." Only as presaging a future attempt to return to bimetallism is there real significance in the present silver policy. Experience has demonstrated that bimetallism at a fixed ratio has little chance of functioning. At the existing legal ratio, nominal bimetallism would instantly become actual silver monometallism. Beyond this further speculation would be futile.

Political significance of silver. Without doubt what has brought silver to its present position of prominence is the strong political pressure exerted by inflationary sentiment and the interests of the silver producers. These are the same forces which fought for bimetallism and secured an artificial increase in the country's silver money between 1875 and 1900. There is little difference in spirit between the present silver policy and the Bland-Allison and Sherman silver purchase acts of 1878 and 1890 respectively.

It is noteworthy that the recent advocates of the policy of raising the price of silver argued that such action would confer great benefit upon any country where silver was still important as a standard of

value. At the time when the silver policy was being formed China was still on the silver standard. Of course the effect of the increase in the price of silver, which occurred in 1934 and subsequently, was profoundly unfortunate for this and other "silver areas." As the value of the Chinese currency in terms of other currencies appreciated, Chinese exports were curbed and imports from foreign nations began to compete more and more severely with domestic products. Thus the appreciation of the Chinese currency in terms of others had a severely deflationary effect upon China, although it may have been of some temporary benefit to those nations whose exports to China were increased. Ultimately, as a result of this pressure, China abandoned the silver standard altogether in 1935, so that the net effect of the silver policy upon China was first to do considerable injury and finally to force the abandonment of the silver standard in the most important area where it still held sway.

The silver policy is really advocated because silver producers in this country and elsewhere are anxious to receive a government subsidy. It is hard to see any valid ground for such favor. Silver mining is one of the world's smallest industries, with an importance about one-sixth that of copper, one-tenth that of sugar, one-sixteenth of cotton, and one-thirtieth of wheat. The value of silver in terms of gold had indeed fallen considerably, but no more than that of many other commodities. As troubled industries are treated with government aid, silver might claim its turn, but there was no reason to give it a position of priority.

Summary. The present American monetary system. The gold reserve act and following executive order completed the process by which the United States definitely abandoned the gold standard and went onto a new money system. While under our definitions this must be classified among the irredeemable paper money systems, it is obvious that the present American system is in certain important respects different from such irredeemable money systems as prevailed during the Revolution and the greenback period. Subject to certain formalities and at the option of the Secretary of the Treasury, gold may be obtained in return for paper money for export and for domestic use commercially and in the arts. In recognition of the former right, the present system has been called

by certain authorities an "international gold bullion standard," a title obviously only partially appropriate. The restrictions which are now contained in the monetary system of the United States were designed primarily to make impossible the hoarding of gold, while not interfering unduly with its use for paying foreign balances or in the arts. This involves denial of free coinage and the use of gold coin in circulation and withdraws the essentials of a standard money system; *i.e.*, free redeemability of all forms of paper money in gold coin (or gold bullion or gold exchange) at the pleasure of the holder and a free gold market, both domestic and for export.

The United States Treasury is obliged to purchase gold bullion whenever offered at the rate of \$35.00 an ounce. This of course prevents the value of the paper money from ever rising above that ratio in terms of gold. On the other hand, the right of the citizen to receive gold bullion for export in exchange for paper money at \$35.00 an ounce will keep the money from falling below that ratio in the international market so long as the Secretary of the Treasury, in the exercise of his discretion, permits such redemption without undue restriction. Exchange of paper money for gold for domestic use in the arts, also limited and at the discretion of the Secretary of the Treasury, is clearly not a sufficient redeemability to assure the gold value of the paper money internally, though it does maintain a certain limited relation between the paper money and gold. The present American monetary system is thus not turned entirely adrift from gold as in the extreme forms of irredeemable paper money systems.

The tables on pages 537 and 538 show the composition of the United States monetary system as it was on June 30, 1932, and is at present (January 31, 1939). Comparison discloses certain important changes. The dollar value of the gold holdings of the Treasury is almost five times what it was (partly due to reduction of the unit). This is all in the form of bullion. There is no longer any gold coin in circulation. The amount of gold certificates has increased greatly, but these are mostly held by the Treasury and the federal reserve banks and agents; only 75 million dollars are still in circulation. The principal part of the gold certificates in the Treasury is due and payable to the federal reserve banks. More

CIRCULATION STATEMENT OF UNITED STATES MONEY, JUNE 30, 1932

Kind of Money	Held in U. S. Treasury	Held outside U. S. Treasury			Total Stock
		By Federal Reserve Banks and Agents	In Circulation		
			Amount	Per Capita	
1. Gold coin and bullion	2,958,560,679	507,272,104	452,732,400	3.62	3,918,565,183
2. Standard silver dollars	501,022,745	8,870,620	30,064,770	.24	539,958,135
3. Token coins :					
<i>a.</i> Subsidiary silver	8,490,556	40,172,383	256,219,820	2.05	304,882,759
<i>b.</i> Minor coin	4,755,771	8,118,336	113,618,813	91	126,492,920
Total coin and bullion	3,472,829,751	564,433,443	852,635,803	6.82	4,889,898,997
4. Gold certificates	775,015,730	715,683,239	5.72	1,490,698,969 *
5. Silver certificates	134,611,482	352,604,719	2.82	487,216,201 *
Total certificates	909,627,212	1,068,287,958	8.54	1,977,915,170 *
6. U. S. notes	2,279,960	55,325,512	289,075,544	2.31	346,681,016
7. Treasury notes of 1890	1,222,150	.01	1,222,150 *
8. National bank notes	16,578,916	19,201,442	700,893,855	5.61	736,674,213
9. Federal reserve bank notes	26,298	100	2,745,642	.02	2,772,040
10. Federal reserve notes	1,406,880	246,760,965	2,780,229,370	22.24	3,028,397,215
Total notes	20,292,054	321,288,019	3,774,166,561	30.19	4,115,746,634
Total monetary stock	3,493,121,805	1,795,348,674	5,695,090,322	45.55	9,004,423,481

* Not included in the total of this column, in order to avoid double counting.

CIRCULATION STATEMENT OF UNITED STATES MONEY, JANUARY 31, 1939

Kind of Money	Held in U. S. Treasury	Held outside U. S. Treasury			Total Stock
		BY FEDERAL RESERVE BANKS AND AGENTS	IN CIRCULATION		
			Amount	Per Capita	
1. Gold	14,681,662,807	14,681,662,807
2. Standard silver dollars	502,735,524	3,426,001	40,917,395	0.31	547,078,920
3. Silver bullion	1,157,793,608	1,157,793,608
4. Token coins					
<i>a.</i> Subsidiary silver	4,081,055	23,895,345	348,140,232	2.66	376,116,632
<i>b.</i> Minor coin	3,637,517	6,405,626	149,345,465	1.14	159,388,608
Total coin and bullion	16,349,910,511	33,726,972	538,403,092	4.11	16,922,040,575
5. Gold certificates	9,132,216,824*	2,815,444,500	74,592,149	0.57	12,022,253,473*
6. Silver certificates	299,809,846	1,268,603,802	9.70	1,568,413,648*
Total certificates	9,132,216,824*	3,115,254,346	1,343,195,951	10.27	13,590,667,121*
7. U. S. Notes	3,617,396	98,892,126	244,171,494	1.87	346,681,016
8. Treasury notes of 1890	1,167,672	0.01	1,167,672*
9. National bank notes	1,367,950	1,980,170	197,536,657	1.51	200,884,777
10. Federal reserve bank notes	502,953	175,050	27,168,873	0.21	27,846,876
11. Federal reserve notes	12,658,165	365,995,330	4,301,229,565	32.89	4,679,883,060
Total notes	18,146,464	467,042,676	4,771,274,261	36.49	5,255,295,729
Total monetary stock	16,368,056,975	3,616,023,994	6,652,873,304	50.87	22,177,336,304

* Not included in the total of this column, in order to avoid double counting.

than a billion dollars' worth of silver bullion has been acquired by the Treasury. No substantial change has taken place in the silver dollars, but the stock of silver certificates has more than trebled. Silver certificates are redeemable in silver dollars, and the Treasury is required to maintain the value of the silver dollars at a parity with other kinds of money, as before. No revolutionary changes have occurred in the other elements of the monetary system.

The currency system of course includes as always individual bank deposits subject to check. Deposits are payable in lawful money, all forms of money being now legal tender.

The irredeemable foreign exchange standard. To complete the analysis of irredeemable money systems, it should be noted that a country may have as the basis of its monetary system foreign exchange upon another country which is not itself on the gold standard. The currency of the first country can of course be no better than that of the second and must therefore be classified as irredeemable. In other respects the principles of the irredeemable foreign exchange standard are similar to those of the gold exchange standard as noted in the previous chapter. A considerable number of nations at present find themselves on the irredeemable foreign exchange standard.

XXIV

THE PRINCIPLES OF BANKING

Financing production. Let us for the moment assume a very simple type of productive organization. There are only manufacturers, wholesalers, retailers, and consumers. Goods pass in a direct line from the manufacturer to the wholesaler, to the retailer, and to the consumer. There are no banks or other financial houses to assist any one of them. When the manufacturer buys materials and pays wages he must pay in cash. When the wholesaler buys goods from the manufacturer he pays cash for them; the retailer and the consumer do likewise. The financial functions of these seem slight, but are they? The manufacturer is compelled to use his own funds while he is manufacturing the goods; the wholesaler ties up his resources in a stock of goods until he has disposed of them to the retailer; and the retailer is in the same position. Even the consumer may buy somewhat in excess of his immediate needs. As a matter of fact, each one is *financing* himself.

Some of the consumers, let us assume, find it impossible to pay for the goods immediately, and they induce the retailer to "trust" them for small amounts. But as a result of this situation the retailer may find himself unable to pay cash for the goods he buys from the wholesaler. He is confident that, if he were given three months in which to make payment, he could sell a certain amount of goods and secure payment from his customers, and so he suggests to the wholesaler that he be allowed to purchase the goods on the basis of deferred payments — that he be given three months' credit. If the wholesaler agrees to this, he assumes the burden of financing the retailer as well as himself. Obviously to do this he must have fairly large resources or be compelled to curtail the extent of his operations, provided he still pays the manufacturer cash on delivery of goods. But he may make to the manufacturer the same plea which the retailer made to him, urging in extenuation his treatment of the retailer. If the manufacturer accedes to his request, he is

financing himself, the wholesaler, the retailer, and the consumer. He is selling goods and receiving in exchange a written or verbal promise to pay for them sometime in the future.

Credit. Here we see the essence of the credit system. Credit involves the purchase of something in the present — merchandise, money, or services — agreeing to pay for it in the future. In business the amount and time of the future payment is usually determined, and the payment is almost always made in money or some other form of currency. But this is not necessarily so. A farmer may borrow a few bushels of seed wheat promising to return a certain number of bushels of wheat when his crop is harvested. This is as truly a credit transaction as when a banker loans \$1,000 to be repaid in money one month from the date of the loan, even though the repayment in the former case is to be in wheat and at a time which cannot be fixed exactly when the loan is made.

The credit system rests ultimately on a belief in the ability of the borrower or purchaser to make the future payment to which he has pledged himself, but its existence is conditioned on the presence of individuals or of a class in society who are willing to advance present goods in return for a promise to pay in the future. It therefore depends on the existence of a surplus, over and above present needs, which can be placed at the disposal of others, and the extent of the credit system is determined by the volume of savings.

In our present example credit has been entirely of the type we shall call *commercial credit*; that is, it has been extended by the manufacturer, the wholesaler, or the retail merchant. This held true in the main of the industry and commerce of former times. The financial houses devoted exclusively or primarily to financing business men are of fairly recent origin. For although it is true that banks have existed for many centuries — the Bank of Venice was founded in the twelfth century — and that we find money lenders and financial houses of importance in the medieval period, they were the exceptions to the general rule, and the producers were forced to rely on their own or their families' resources or on each other. It was not until the eighteenth and nineteenth centuries that banks and other institutions so increased in number and resources as to become a vital part of the economic fabric and *bank credit*

(a promise by a bank to pay) surpassed in importance what we have here called commercial credit (a promise by a manufacturer, wholesaler, or retailer to pay).

The commercial bank. During the history of banking there have developed several different kinds of banks, and at the present day we need to distinguish at least three types; namely, commercial banks, trust companies, and savings banks. Land banks might be added as a fourth class. It is however the commercial banks which play the predominant rôle in the modern economic organization, and it is to this class of banks that our attention will be chiefly directed. A few words at the close of this chapter will suffice for the other types.

The reader may be presumed now to know that a commercial bank's business consists primarily in making loans to and receiving deposits from its customers. Let us plunge at once into the middle of our subject by assuming a moderate-sized city bank and following through in detail from beginning to end some of its typical operations.

Lending: Example of borrowing. Let us assume that J. C. Miller, a dry goods merchant, is engaged in buying his stock of goods for the early spring trade. As is usual, he does not have enough capital of his own for the purchase of his entire stock but depends regularly upon the financial assistance of his bank. He therefore goes to the bank and asks to borrow, let us say, \$50,000. The officers of the bank, after satisfying themselves as to his character and the state of his business, agree to accept his promissory note for \$50,000 payable, let us say, sixty days from date. The note would be in some such form as the following :

\$50,000

New Haven, Conn., Aug. 25, 1939

Sixty days after date I promise to pay to the order of the City Bank fifty thousand dollars. Value received.

(Signed) J. C. MILLER

Interest and discount. The compensation which the bank gets for supplying Mr. Miller with funds is called interest or discount,

and the process is called loaning or discounting, depending upon the way the note is drawn. If the note is drawn *with interest*; that is, if Mr. Miller promises to repay at maturity the \$50,000 plus interest at 6 per cent (the rate asked by the bank) for 60 days, he is said to be getting a *loan* of \$50,000 and to be paying *interest* on \$50,000. In this case, at the time of borrowing he will receive \$50,000 and at maturity will pay \$50,500. The bank properly regards the promissory note as worth \$50,000 at the time of borrowing and so enters it among its assets.

If however Mr. Miller writes his note for \$50,000 *without interest*, the note is not worth \$50,000 this day to the bank, and it must determine what value to allow him for it. The bank knows that 60 days from date it will bring \$50,000; according to the principles of valuation with which the reader is now familiar,¹ it might be supposed that the bank would determine what sum put at interest for 60 days at 6 per cent would amount to \$50,000; namely, $\$50,000 \div 1.01$, which is \$49,504.95. The \$495.05 represents the *true discount*, or the amount by which the face value exceeds the present value of the note.

Banks are accustomed however, because of the extra profit and the easier computation, simply to deduct 60 days' interest at 6 per cent on the \$50,000 face value, leaving \$49,500. In banking parlance the \$500 is called *bank discount* or simply *discount*, and the \$49,500 is called the *proceeds*. Mr. Miller in this case really procures only \$49,500 from the bank but pays interest on \$50,000; and if he had needed exactly \$50,000 it would have been necessary for him to draw his note for \$50,505.05. Furthermore when the bank, upon receiving the note, enters it among its assets at \$50,000, as is the custom of banks, it overvalues the note by \$500 according to its own computation; the discount will be earned only as the note appreciates in value to \$50,000 at maturity.

To make the picture complete, we may suppose that the bank is asked to discount a note for \$50,000, bearing interest at 4 per cent, and maturing in 60 days. If the bank's rate of discount is still 6 per cent, the bank clearly would be unwilling to give the borrower the full face value of the note. At the end of 60 days the note will be

¹ See Chapter XV.

worth \$50,000 plus the interest at 4 per cent for 60 days (\$333.33) or a total of \$50,333.33. Discounting this amount at 6 per cent for 60 days, we find that the discount is \$503.33 and the proceeds therefore \$49,830.00.

The term *discount* is used both to refer to the interest deduction and to describe the whole process of evaluating and disposing of the note. In the financing of commerce the bank discount method is more commonly used at the banks, but there are many types of financing in which the loan method prevails.¹

Proceeds. The next question is as to the form in which Mr. Miller shall receive the funds borrowed. There are ordinarily three forms, (1) money (other than the bank's own notes), (2) the notes of the bank, (3) deposits. The first is simple enough and needs no explanation. In the second case the bank gives Mr. Miller, in exchange for his promissory note, its own promissory notes. It is a swapping of notes. But the notes of the bank are money; they are generally acceptable throughout the community, and Mr. Miller has received what he wanted, the means for making immediate payment for his goods. Indeed the ordinary business man does not distinguish between different kinds of money; to him there is practically no distinction between the first two cases. To the bank, as will appear later, there is a very real distinction. Both of these forms of payment however are the exceptional cases, at least in the United States. In the great majority of instances the borrower receives the proceeds of his note, not in money ("cash"), but in the form of a deposit credit.

When money is taken to a bank and left — "deposited" — the bank credits the customer's account on its books, and the latter obtains the right to withdraw the specified amount of money, or any part of it, at will. Now when a note is discounted the amount of the proceeds is (in the ordinary case) credited to the account of the borrower as a deposit, precisely as though he had brought money. If it is any aid to visualizing the process, it may be imagined that the customer received the proceeds in money and immediately stepped over to the receiving teller's window and deposited the

¹ As is the case with most economic categories, the terms, *loan* and *discount*, *interest* and *discount*, are loosely and interchangeably used by the public.

money. Actually however no money has been involved; this is indeed not a money transaction. It is once more a swapping of credit. The bank receives the customer's promissory note and gives him in exchange the right to receive money from it on demand, which is precisely what he would have if he had taken the bank's notes instead.

Principal types of trade paper. We must now retrace our steps in order to consider the several forms of paper which may be used in the ordinary commercial borrowing at a bank. There is first the simple promissory note of the borrower. Such a note may be *unsecured*, or it may be *secured* by "collateral security." In the latter case the borrower delivers to the bank along with the note certain property of his, worth at least as much as the amount of the note, with an agreement that in case the note is not paid when due the bank may sell the security and so reimburse itself for the note. Stocks and bonds of corporations and government bonds are the property most used as collateral security, though warehouse receipts for certain commodities, bills of lading, etc., may be used. The secured note is one of the common forms of paper discounted by banks.

In the example which we have used, Mr. Miller discounted at the bank a note, either secured or unsecured, and obtained the \$49,500 which he used to pay the wholesaler. The purchase of his stock in trade might however have been somewhat differently financed. It is common practice for the wholesalers to sell to their customers on credit; that is, to accept their promissory notes in place of cash. Suppose that Mr. Miller has such an arrangement with a wholesaler, R. B. Norton, in the same city. He will give Mr. Norton in exchange for the goods his promissory note for \$50,000 payable in 60 days. The wholesaler, Norton, being himself actively engaged in business with all his capital, will not be willing to hold Miller's note till maturity; he needs the cash before that. Norton therefore avails himself of the services of the bank. He first *endorses* Miller's note; that is, signs his own name, thereby making himself liable for its payment in case Miller should not pay. He then takes the note to the bank, which discounts it precisely as in the case when Miller borrowed on his own note. The proceeds are given to Norton, either in money or, more likely, as a credit on his deposit account.

The note becomes the property of the bank, and when it is due Miller will pay, not Norton, to whom he gave the note originally, but the bank. A note thus endorsed is called "two-name paper" as compared with the simple promissory note of the maker, called "one-name paper."

The paper offered to the bank may take still a third form; namely, the *draft* or *acceptance*. The retailer, instead of giving his promissory note in return for goods received, may give the wholesaler the authority to *draw* upon him. This, in the case we are using for illustration, would mean that when Norton sent the goods to Miller he would write out an order upon Miller directing him to pay the \$50,000 in sixty days to Norton's bank. This draft would then be taken to Miller, who would accept it; that is, write the word "accepted" and his signature across its face, which is his acknowledgment of his obligation to pay at the maturity of the draft. This draft would be in form similar to the following:

	New Haven, Conn., Aug. 25, 1939
To J. C. Miller	
On October 25, 1939, pay to the order of the City Bank fifty thousand dollars (\$50,000).	
The transaction which gives rise to this instrument is the purchase of goods by the acceptor from the drawer.	
Accepted at New Haven, Conn.	
Aug. 25, 1939	
Payable at the City Bank,	
New Haven, Conn.	
(signed) J. C. MILLER	(signed) R. B. NORTON

A draft when thus accepted is called an *acceptance*. It is two-name paper since the drawer, Norton, is liable as well as Miller, upon whom it is drawn. Mr. Norton then takes the acceptance to the bank and discounts it precisely as has been explained in the case of Miller's promissory note discounted by Norton. At the maturity of the draft, Miller will pay the bank. These various forms of business men's notes and drafts are all called *trade paper*.

A bank's balance sheet. These matters may be made clearer and the foundation laid for our further investigation, if we now give

some attention to the balance sheet of the bank. A commercial bank is ordinarily a corporation, and to it may be applied what has already been learned about the organization and accounts of corporations. Avoiding details not here essential, we may say that a bank is organized about as any other business corporation.¹ The first balance sheet might read somewhat as follows :

STATEMENT OF THE CITY BANK, AUGUST 1, 1939

<i>Assets</i>		<i>Liabilities</i>	
Cash	\$400,000	Capital stock	\$400,000

The bank proceeds at once to buy land and a banking building, furniture and supplies, etc., for \$60,000, and thereafter its balance sheet reads as follows :

STATEMENT OF THE CITY BANK, AUGUST 10, 1939

<i>Assets</i>		<i>Liabilities</i>	
Real estate	\$ 45,000	Capital stock	\$400,000
Furniture, etc.	15,000		
Cash	340,000		
	<u>\$400,000</u>		<u>\$400,000</u>

Discounting a note. Now comes Mr. Miller, the merchant, with his request for a loan. As has been already assumed, he offers his promissory note for \$50,000 payable in sixty days, the bank's rate of discount is six per cent, the discount is \$500, and the proceeds, \$49,500. Mr. Miller elects to receive "cash" other than the bank's own notes. The effect of this transaction upon the bank's condition is shown in the following statement :

STATEMENT OF THE CITY BANK, AUGUST 25, 1939

<i>Assets</i>		<i>Liabilities</i>	
Loans and discounts	\$ 50,000	Capital stock	\$400,000
Real estate	45,000	Undivided profits	500
Furniture, etc.	15,000		
Cash	290,500		
	<u>\$400,500</u>		<u>\$400,500</u>

It will be noted that the \$50,000 promissory note acquired by the bank is listed among its assets opposite the heading "loans and

¹ See Chapter V.

discounts," that the cash has been reduced by the payment of \$49,500 to Mr. Miller, and that the difference between these amounts, the discount, appears as undivided profits. As has been noted, these are fictitious profits and represent simply an over-valuation of the note; at maturity, if the note is paid, they will be true profits. Obviously the effect on the bank's statement will be the same whether Mr. Miller borrows by means of an unsecured note or a note secured by collateral. Even in case he finances his business by giving his promissory note to the wholesaler or permitting the wholesaler to draw upon him, the final result is the same, the only difference being that in these cases (that is, the endorsed note or the acceptance) Miller's obligation is discounted at the bank by Norton, the wholesaler, instead of by Miller himself. In any case the bank comes into possession of Miller's obligation, which he must pay to the bank at maturity.

The statement will however differ according to the form in which the proceeds are given to the customer. The statement above illustrates the case where the proceeds are taken in the form of money other than the bank's own notes. If the bank had given its notes the statement would have read thus:

STATEMENT OF THE CITY BANK, AUGUST 25, 1939

<i>Assets</i>		<i>Liabilities</i>	
Loans and discounts	\$ 50,000	Capital stock	\$400,000
Real estate	45,000	Undivided profits	500
Furniture, etc.	15,000	Notes	49,500
Cash	340,000		
	<u>\$450,000</u>		<u>\$450,000</u>

If finally the proceeds had been credited to the deposit account of the borrower, the statement would have been as follows:

STATEMENT OF THE CITY BANK, AUGUST 25, 1939

<i>Assets</i>		<i>Liabilities</i>	
Loans and discounts	\$ 50,000	Capital stock	\$400,000
Real estate	45,000	Undivided profits	500
Furniture, etc.	15,000	Deposits	49,500
Cash	340,000		
	<u>\$450,000</u>		<u>\$450,000</u>

A normal bank statement. To carry our illustration somewhat further, let us assume that during the next two months the bank has done a flourishing business, making loans, discounting paper, and correspondingly increasing its deposits and notes and profits, that a part of the latter has been transferred to surplus, and that the bank has made some investments in securities. The statement, corresponding more closely to a normal banking business, may be assumed to be as follows:

STATEMENT OF THE CITY BANK, OCTOBER 25, 1939

<i>Assets</i>		<i>Liabilities</i>	
Loans and discounts	\$1,250,000	Capital stock	\$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	12,327
Furniture, etc.	15,000	Notes	50,000
Reserve	182,460	Deposits	1,405,133
	<u>\$1,942,460</u>		<u>\$1,942,460</u>

The reader will note the appearance in this statement of a new term, the *reserve*. This is an inclusive term indicating the sum of such assets as are available for payment of the bank's liabilities, especially its notes and deposits. The term is usually synonymous with cash or items equivalent to cash. Banks in the United States do not generally use this term in their statements, specifying rather the separate items which compose their legal reserves.

Bank deposits: Definition. Attention has thus far been directed principally to the function of discount. We must now inquire further into the subject of deposit. *A bank deposit is a right to receive money from a bank, evidenced by an entry on the bank's books and by the customer's passbook, duplicate deposit slip, or other device.* Deposits are classified as demand or time, the bank being obligated to pay the former on demand of the depositor but obligated to pay the latter only after a certain number of days' notice.¹ Deposits may result from the discount of trade paper as illustrated above, or from deposit of money, or from deposit of checks drawn by another depositor, either of this or another bank, or in certain other ways that need not concern us here.

¹ Thirty days' notice is commonly required.

To illustrate the deposit of money, let us now suppose that a customer brings in \$10,000 of money for deposit. The bank's statement immediately thereafter will read thus:

STATEMENT OF THE CITY BANK, OCTOBER 25, 1939

<i>Assets</i>		<i>Liabilities</i>	
Loans and discounts	\$1,250,000	Capital stock	\$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	12,327
Furniture, etc.	15,000	Notes	50,000
Reserve	192,460	Deposits	1,415,133
	<u>\$1,952,460</u>		<u>\$1,952,460</u>

It will be noted that nothing on the bank's statement distinguishes between deposits obtained by paying money to the bank and deposits obtained by the process of discount. There is no difference. What the depositor has is exactly the same in either case, not money, but the right to receive money. Deposits are not, as is sometimes popularly supposed, "money in the bank." Banks never have on hand a stock of money equal to their deposits. That could hardly be expected, since, as we have seen, deposits may originate without any money being given to the bank. Even when money is deposited, the money ceases from that moment to be the property of the depositor; it belongs to the bank and may be spent by the bank as it sees fit. The depositor has simply the right to receive a certain amount of money whenever he chooses to demand it of the bank.

Discount and deposit. In fact both discount and deposit, contrary to the meanings popularly attached to the terms "loan" and "deposit," are actually exchanges, transfers of certain rights for other rights or for money. When the bank discounted Mr. Miller's promissory note it acquired a property right against Mr. Miller and in exchange gave him a property right against the bank; namely, a deposit. Quite properly the bank lists the note as an asset and the deposit as a liability. On the other hand Mr. Miller enters the note on his accounts as a liability and the deposit at the bank as an asset. When Mr. Norton, having accepted Mr. Miller's promissory note in payment for goods, discounted the note, he sold it to the bank in exchange for a deposit, or perhaps for bank notes or some other

form of money. Money deposited at the bank is, so to speak, "spent" — paid for a deposit.

The double function of discount and deposit is ordinarily a swapping of credit — the person's promissory note or draft or acceptance for the bank's credit in the form of notes or deposits. If it be asked wherein lies the advantage of merely exchanging one form of credit for another, the answer is that what the bank's customer wants is a medium for making payments. His own promissory note will not be generally accepted; the bank's note will. He has exchanged a form of credit not generally known and acceptable for the well-known and acceptable credit of the bank. The same is true of deposits, since practically all the customer's payments may readily be made by check. Herein is the fundamental characteristic of bank credit and of the service rendered by the commercial bank.

The check: Definition. The right to receive money which a bank deposit gives the depositor is exercised by means of the check. If Mr. Miller desires to withdraw fifty dollars from the bank in order to make a certain payment or for pocket money, he makes out an order upon the bank; *i.e.*, a check, in the following form :

No. 731	
New Haven, Conn., Oct. 25, 1939	
THE CITY BANK	
Pay to the order of.....	<i>Myself</i>
.....	<i>Fifty</i>Dollars.
	\$50.00
	(signed) J. C. MILLER

He presents this order to the bank, which thereupon gives him fifty dollars in money and deducts fifty dollars from the amount due him on its books. It will also deduct fifty dollars from its reserve item in the assets column.

If Mr. Miller wished to pay a doctor's bill of fifty dollars, he might obtain from the bank fifty dollars by means of a check like the above and then take the money to the physician. But a simpler

method would be to order the bank to make the payment direct to the physician. For this purpose a check would be used in this form :

No. 732

New Haven, Conn., Oct. 25, 1939

THE CITY BANK

Pay to the order of	Robert Brown.....		
.....	Fifty..	Dollars.	\$50.00
(signed) J. C. MILLER			

From these examples the nature of the check is apparent. *A check is a written order by a depositor directing his bank to pay money to some person.* The person to whom payment is to be made may be the depositor himself or the bank or any specified third party or the bearer. Checks are usually made out on printed forms like those in the illustration above, furnished by the bank, but such a printed form is not necessary.

Payment by check: Within the bank. Let us follow the history of this check. The check is handed or sent by mail to Dr. Brown, who takes it to the bank and demands the money. Certain essential facts must be established before the paying teller of the bank will hand over the money. First he must be satisfied as to the genuineness of the check, which is usually sufficiently attested by the signature of the maker, with which the teller is familiar. Next he must be certain that Mr. Miller actually has as much as fifty dollars to his credit on the bank's books. The teller must also be certain that it is Dr. Brown himself who is asking for payment; if he is not personally known he will have to be identified. Before payment is made the check must be endorsed; that is, the payee, Robert Brown in this case, must sign his name on the back of the check. Endorsement is the payee's acknowledgment that he has received the money; it makes the check evidence of payment for the bank and for the maker. Having made the payment, the bank debits the amount to Mr. Miller's account, and eventually, when the account

is balanced, the cancelled check will be returned to him to be kept as evidence of the payment. The effect of this operation will be to alter two items on the bank's balance sheet, making it read as compared with the last statement (on page 550) as follows:

STATEMENT OF THE CITY BANK, OCTOBER 26, 1939

<i>Assets</i>		<i>Liabilities</i>	
Loans and discounts	\$1,250,000	Capital stock	\$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	12,327
Furniture, etc.	15,000	Notes	50,000
Reserve	192,410	Deposits	1,415,083
	<u>\$1,952,410</u>		<u>\$1,952,410</u>

Instead of demanding the money, the payee may "deposit" the check. He endorses the check and takes it or sends it by mail or messenger to the bank, with the request that the amount stated be credited to his deposit account. The bank must be certain as to all the facts except identification, just as when money is demanded, and must be satisfied as to the genuineness of the endorsement. It will then credit the account of the payee (*e.g.*, Robert Brown) and debit the account of the maker (*e.g.*, J. C. Miller). The result of this transaction is exactly as though the payee had collected the money from the bank and then immediately deposited it. The balance sheet of the bank is not changed at all; no cash has been taken out, and the total amount due depositors is not affected by a transfer from one depositor's account to another's.

The check thus furnishes a means of making payments so convenient, safe, and generally advantageous that, in the Anglo-Saxon countries at least, it is availed of very generally, not only by business men but by others. Deposit accounts are kept at the bank. Money when received is regularly deposited at the bank, and the same is true of the far greater volume of receipts which come in the form of checks. Conversely fully four fifths of all payments are made by means of checks rather than with money. The bank thus becomes an agent for making payments between the people by the transfer of deposits from one to another.

Collection between banks: In general. We have proceeded thus far upon the assumption of one bank with which all persons

concerned have their dealings. As a matter of fact in every place of any size there will generally be found two or more banks. Moreover the useful device of payment by check has not been permitted to remain a merely local convenience but has absorbed the great mass of payments between persons in different cities, states, and even nations. This introduces certain complications, which must now be investigated.

When one receives a check, the obvious thing is to take it to one's own bank and either deposit it or ask for the money. This is done regardless of whether the check happens to be drawn on this same bank or upon another. In the first case, the matter is simply arranged, as has been explained. Even in the second case, the bank, though under no obligation to cash or credit the check, will ordinarily do so as an accommodation to its customer, usually making a small charge for performing this service. Having given its customer the cash or credited his account, the bank thereupon becomes possessed of the check with its claim upon the other bank. If the other bank is in the same place, the check will ordinarily be presented and paid within twenty-four hours. A longer time may be required when the debtor bank is in another city. The other bank of course then debits the account of its depositor who drew the check.

The first bank, in performing this service, accepts at its face value a check drawn by a stranger, the genuineness of whose signature it has no means of testing and the state of whose account in the other bank is wholly unknown to it. For protection it relies simply upon the endorsement of the check by the customer who presented it. His endorsement is his guarantee that the check is good and binds him to reimburse his bank if the check should prove not good. Checks sometimes pass through several hands before finally reaching the bank upon which they are drawn. Each person (individual or bank) who thus passes on the check is ordinarily required to endorse it, and each endorser makes himself responsible for the soundness of the check, unless signed "without recourse."

Thus the check system for payment of money or transfer of deposits begins to operate as soon as there are two or more banks in the town. Daily each bank sends a messenger to each of the

others and collects for all checks of each other bank which it has paid. But in a large city with many banks this business of collecting by messenger becomes troublesome and expensive. In a city with fifteen banks there might have to be 15×14 , or 210, separate calls each day. Sooner or later the point is reached where it is worth while to avoid this great labor and expense by organizing a clearing house.

Clearing house. The clearing house is an association of the banks of a city through which all the claims and obligations between the banks for paying each other's checks and drafts are settled. While the fundamental nature of the clearing house and its business is everywhere the same, there are numerous variations of detail. The following account is based primarily upon the methods employed in the New York Clearing House. The business of clearing takes place daily at certain hours. Before each clearing there occurs in each bank the same preparatory operations. The clerks gather together all checks of the other banks which have been received since the last clearing and arrange them in bundles, one for each bank. A list is made showing the amount thus due from each of the other banks, with the total due from all the banks. Sometimes there is also made out a ticket for each of the other banks showing the amount due from that particular bank. Then two clerks, gathering together all the bundles, the tickets, and the list, proceed to the headquarters of the clearing house. Here there is a room containing a raised desk at one end for the governor and his assistants and on the floor a row or rows of desks, one for each member bank. Of the two clerks one takes his place behind his own bank's desk while the other, carrying the bundles of checks and the tickets, stands in front of the desk. The list showing the amount due this bank from all others has been sent up to the governor's desk. In like manner, each of the other banks is represented by two clerks, one behind and one in front of its own desk.

At the exact moment for beginning the operation of clearing a signal is sounded on the governor's desk. The clerks in front then walk along the row of desks, depositing at each the bundle of checks and the ticket against that particular bank. In a minute or two the checks and tickets have thus all been distributed, and it takes only

a few minutes more for each desk clerk to add up the tickets or the sums on the outside of the envelopes and so ascertain the amount his bank owes to all of the other banks. These results are reported to the governor, whose assistants are able quickly to discover whether any error has been made. As soon as the figures, corrected if necessary, are compiled, the governor announces the result. Each bank now knows exactly how much it owes all the other banks and how much all the other banks owe it. These two sums are regarded respectively as a debt to the clearing house and a credit due from the clearing house. The balance represents an amount which the bank must pay to the clearing house or receive from the clearing house as the case may be. Thus by one payment either to or from the clearing house, each bank settles its accounts with all the other banks. Of course the amount payable by all the debtor banks to the clearing house is exactly equal to the amount payable by the clearing house to all the creditor banks, and the clearing house itself comes out even each day.

Clearing for outside banks. All the banks of a city do not necessarily belong to the clearing house. An outside bank will either make its collections by messenger and make its payments to messengers sent by the other banks in the primitive way, or it may arrange to clear through another bank which belongs to the clearing house. In this case the agent bank presents as its own the claims of the outside bank and accepts as obligations against itself claims of the other clearing house banks against the outside bank.

The volume of clearings. The volume of the business thus simply and smoothly effected by a large city clearing house is enormous. In the New York City Clearing House there were exchanged, during the month of December, 1938, checks and similar obligations to the amount of \$17,950,623,781. For the year 1938, the New York City clearings were over 165 billion dollars. The New York City Clearing House stands far ahead of any other in volume of clearings. For the whole of the United States the clearings during the month of December, 1938, were about $30\frac{1}{2}$ billions, and for the whole year 1938, $293\frac{1}{4}$ billions. Since our available sources of such statistics do not include reports from all the cities having clearing houses, the figures here given for the whole country are materially

less than the actual totals of all clearings. It should also be noted that the year 1938 was not one of active business; at the year's close the country was definitely in a state of industrial depression. In 1929, for example, clearings were about two and a half times those of 1938. One can imagine what a gigantic task would be imposed upon the banks were they still compelled to rely upon the primitive method of exchanging their obligations.

The reader should be reminded that even these figures do not give a true picture of the total volume of banking business. In fact the recent movement toward bank consolidation, while accompanied by increased business, has obviously tended to a diminished use of the resources of the clearing houses. A more reliable indication of banking business is presented by the figures of debit check transactions, which for the year 1937 totalled over 469 billion dollars for the principal cities of the country.

Out-of-town checks. Checks are regularly used for payments in other towns. When goods are purchased by mail order, payment is ordinarily made by a check drawn by the purchaser upon his own bank and sent by mail to the seller, it may be a thousand miles away. The maker of the check gives himself no further concern over the matter. The receiver of the check deposits it in his own bank, receives credit for it, and concerns himself no further. His bank has the task of collecting from the other distant bank. Thousands of payments are thus made every day, and every day the mails carry great numbers of checks back and forth all over the country. There is thus imposed upon the banks a service of great importance and enormous volume. An elaborate system has been developed whereby this business is handled with swiftness and certainty and the minimum of expense and, by offsetting checks against each other, with remarkably little actual transfer of money.

Paying a loan. We must now return to the further study of the business of discount and deposit. We have used for illustration the discount of a note for \$50,000 for sixty days at the City Bank by J. C. Miller. On October 25, 1939, the loan matures and must be paid at the bank. This is equally the case whether Mr. Miller borrowed directly from the bank on his own promissory note or gave his promissory note to the wholesaler, Norton, or accepted a

draft drawn upon him by Norton, so long as the paper has been discounted at the bank.

Mr. Miller now presents himself and offers to pay his note. He may pay in any of five ways: (1) in money, other than the notes of this bank, (2) in notes of this bank, (3) by check upon a deposit account in this bank, (4) by a check upon an account in some other bank, or (5) by a bank draft; *i.e.*, a check drawn by a bank either on itself or on its account in some other bank. The effect upon the bank's condition is readily determined. In any case the loans and discounts will be reduced by \$50,000, representing the note paid and returned to Mr. Miller. If the note is paid in money the bank's cash, or reserve, will rise by \$50,000, and the same result will follow, within a few days at most, if payment is made by a check or a bank draft on another bank, since this will be promptly presented to the other bank, through the clearing house or by mail or messenger, and collected. If the note is paid by giving notes of this bank, the note item on the liability side of the account will be reduced by \$50,000. Finally if payment is by a check either on Mr. Miller's account or on the account of another depositor in this bank, the bank's deposits will be correspondingly reduced. Two changes of \$50,000 each counteract each other, and the statement remains in balance. The reader can easily draw up a new statement representing the bank's condition after payment of the note according to any one of the possible assumptions.

The normal business of banking. Acquaintance has now been made with the more important operations which make up the normal daily business of the ordinary commercial bank. The bank's "portfolio" of loans and discounts contains numerous notes and acceptances of individuals and corporations, to which new notes are being added daily and from which notes are being as constantly withdrawn when they mature and are paid. The bank's deposits are being constantly added to as new loans are made or as customers bring in for deposit either money or checks upon other banks, while on the other hand withdrawals by check are continually tending to reduce deposits, particularly in connection with the payment of loans. The bank's notes outstanding tend likewise to rise and fall in harmony with the loans and discounts. From time to time

securities held by the bank are sold or new ones purchased, with corresponding changes in the "securities" item. In connection with many of these operations changes are continually occurring in the bank's reserve, as cash is either received or paid out. The business is normally profitable, and the "undivided profits" item gradually increases, reduced at intervals by the declaration of dividends or by transfer to surplus.

Interrelation of loans and discounts, deposits and notes, and cash. Whenever a business man borrows from his bank, taking the proceeds in a deposit credit or in notes, the immediate effect is to increase both the assets and the liabilities of the bank by the amount borrowed; loans and discounts increase on the one side, deposits or notes, on the other. Very soon however the borrower will draw checks against his now flush account or transfer his bank notes to other persons. Let us assume for the moment that there is only this one bank. It is conceivable that these other persons to whom the borrower has transferred his claim against the bank may all demand cash. In that case the net result is an increase in the bank's loans and discounts and a corresponding decrease in its cash. But this result is quite improbable. The majority of the newly arisen claims against the bank will be presented to it for deposit, and the final result will be approximately the same as the immediate result; namely, an increase in loans and discounts balanced by an increase in notes or deposits except to the extent that cash has been withdrawn and held by the people as additional pocket money. Even if the original borrower had taken cash, the final result would be about the same, since most of the persons to whom he made payments would doubtless soon deposit the cash at the bank.

Relinquishing now our assumption of only one bank, it is obvious that many of the persons to whom the original borrower makes payment will deposit their claims against his bank in other banks, thus increasing their deposit liabilities. These banks will promptly present their claims against the first bank and demand cash payment. The cash of the first bank will be reduced while that of the other banks will be correspondingly increased, saving again such small amount as may remain outside as increased pocket money of

the people. Taking all the banks as a whole, the net result of this series of operations is exactly the same as when one bank only was concerned; namely, an increase of loans and discounts, balanced by an approximately equal increase of notes and deposits and a relatively small decrease of cash. This result is likely to be accompanied however by a considerable disturbance of the relative cash holdings of the several banks. When one bank of the group makes a loan or discount, its cash will quickly be drawn down by the demands of the other banks to whom the borrower has transferred his claims. It is conceivable that, by the same token, borrowings from the other banks might give this bank claims against them through which its cash would be restored. But this would occur only if there were the same relative expansion of loans among all the banks. Although such expansion is generally rather uniform, there are seasonal fluctuations, during which the banks of one section are drawing heavily on those of another section, and other changes may at times cause long time or permanent modifications in the relative resources of different banks. For any bank in a system the amount of loans and discounts it can make is limited quite closely by the amount of money and immediately collectible items the bank has received from its customers.

In the normal operations of bank credit, increased borrowing results in the expansion of both loans and discounts and notes and deposits. Conversely when borrowing slackens and notes are being paid off, there follows a decrease in these major items of the banks' balance sheets. In general loans and discounts on the asset side and notes and deposits on the liability side tend to rise and fall together with expanding and contracting business.

The reserve: Its function. The deposits and notes of a bank represent claims against the bank held by outsiders and payable in cash on demand. The bank must be ready to pay any note holder or depositor any amount to which he is entitled at any time; refusal to pay means "suspension" or failure. For this purpose the bank must have always on hand a sufficient stock of cash, appropriately called the reserve. No commercial bank however ever holds a reserve equal to the sum of its demand liabilities. In fact the reserve will seldom be as much as twenty-five per cent and may be

much less. In the last illustrative statement (page 553) the bank's liabilities payable on demand (notes and deposits) appear as \$1,465,083, for which there is a reserve of \$192,410, a *reserve ratio* of a little more than thirteen per cent. Evidently a simultaneous demand from all depositors and note holders would wreck the bank. But of course no such simultaneous desire for payment is to be expected, and the bankers learn by experience about what demands may be anticipated from day to day. They make provision for a corresponding reserve with a proper margin for emergencies, and it is thus that the reserve ratio is determined except when legal requirements interfere.

Profits vs. safety. The banker is impelled by two opposing motives — profits and safety. The bank derives its profits principally from the making of loans and discounts, and the larger its portfolio of loans and discounts the larger in general will its profits be. But as loans and discounts are made cash is immediately withdrawn, or deposits are created, or the bank's note issues are increased. Thus the reserve ratio falls, and the bank's condition becomes proportionately less safe. To the banker's desire for profits is thus opposed the necessity of keeping a safe ratio between the reserve and the demand liabilities. The problem is solved differently according to the circumstances and the policy of the particular bank. From the point of view of assets, if the bank's clientele consists mainly of conservative business men engaged in sound and safe enterprises, a relatively low ratio may be ample. If the bank is loaning for more venturesome enterprises, it must safeguard itself by a larger reserve ratio. From the point of view of liabilities, the minimum reserve that is still safe is affected by the number and character of the depositors of the bank, the state of public confidence, and the stability of the industry of the community. A bank having few depositors, with large accounts each, may well keep a larger reserve than a bank with thousands of depositors with small balances. When the depositors are largely foreigners, suspicious and not well acquainted with banking, or when the state is threatened with war, the stock and produce markets shaken, or the local community rocked by business failures or other disturbing events, or when the local industries are subject to capricious stops and

starts as in the oil fields, or when the community is dependent upon one industry alone, it is well for banks to carry heavy reserves. Some bankers are by nature venturesome, "progressive," or reckless; they expand their business at the risk of safety. Others are cautious and "conservative" and keep their reserve ratios well above the danger line.

How the reserve ratio is controlled. There are five principal means of controlling the reserve ratio : (1) by the purchase and sale of securities, (2) by rediscounting commercial paper, (3) by borrowing from another bank, (4) by calling in loans which are payable on demand, and (5) by making it more or less easy or advantageous to borrow. The latter may be accomplished by changing the rate of discount, by simple refusal to loan, by moral suasion inducing applicants not to apply for loans, by raising or lowering the collateral security requirements, or by requiring more or fewer names on the paper.

Buying and selling securities. As we have seen, a bank usually invests a part of its resources in corporate securities, the main purpose being to derive some income upon funds which if kept as idle cash in the reserve would be earning nothing, while at the same time having something which may be converted into cash whenever needed. The bank therefore chooses for investment securities which may be sold at short notice. In the statement on page 553 the bank has a reserve ratio of a little more than thirteen per cent. Let us assume that this is considered too small and that securities carried on the books at \$100,000 are sold for that amount, reducing the stocks and bonds to \$350,000 and increasing the reserve to \$292,410.

The reserve ratio is now $\frac{292,410}{1,465,083}$, or almost exactly twenty per

cent. The bank has sacrificed the future interest or dividends on \$100,000 of securities, but it has made its position materially safer. Conversely a purchase of securities would reduce the reserve and consequently the reserve ratio. It is because of their availability thus to strengthen the reserve that a bank's readily convertible securities are sometimes called its "secondary reserve," though this figurative use of the term must not lead one to think that securities can ever be real reserve.

Rediscounting and borrowing. Instead of selling securities, the bank may strengthen its reserve ratio by selling some of the paper in its portfolio. This is called *rediscounting*, which means to discount again. Some investor is found, either another bank or a broker, willing to take certain promissory notes or drafts at a price determined by discounting them at a certain rate for the time that must still elapse before their maturity. Suppose our bank selects from its portfolio a note for \$60,000, originally acquired two months ago and payable three months from its date, having therefore still one month to run. If another bank is ready to rediscount it at eight per cent, the City Bank will receive for it \$59,600; that is, the face value less one month's interest at eight per cent. The result of this operation, starting from the statement on page 553, will be reflected in a new statement, as follows:

STATEMENT OF THE CITY BANK, OCTOBER 27, 1939

<i>Assets</i>		<i>Liabilities</i>	
Loans and discounts	\$1,190,000	Capital stock	\$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	11,927
Furniture, etc.	15,000	Notes	50,000
Reserve	252,010	Deposits	1,415,083
	<u>\$1,952,010</u>		<u>\$1,952,010</u>

It will be noted that loans and discounts have been decreased by \$60,000, the reserve is increased by \$59,600, while the discount is reflected as a loss, decreasing the undivided profits by \$400. The reserve ratio is now $\frac{252,010}{1,465,083}$, or something more than 17 per cent, instead of 13 per cent as it was before, showing that the desired result has been accomplished.

Similarly a bank may be able to obtain cash by borrowing from another bank on its secured or unsecured short-term promissory note. Referring to the statement on page 553, we see that if \$50,000 in cash were secured in this way the reserve would be increased to \$242,410, and the reserve ratio would be $\frac{242,410}{1,465,083}$, or something over 16 per cent, as compared with 13 per cent.

Calling loans. In contrast to the loans which have a definite date of maturity, there are loans which are made payable on demand.

If a bank chooses to call for the immediate payment of such loans it can increase its reserve ratio. If payment is made in cash or in checks on another bank the reserve will increase, while payment in notes of this bank or in checks on this bank will decrease notes and deposits; in either case the reserve ratio rises.

Control by the rate of discount. By the methods thus far described the banker is able to make a quick change in his reserve ratio. Another method of regulation is by varying the rate of discount. As has been observed, in the course of a bank's daily business new loans are constantly being made and notes held are constantly maturing and being paid. There is thus an inflow and an outflow of paper, which under ordinary circumstances may about offset each other, keeping the amount of loans and discounts fairly constant though the individual notes held are always changing. Now the payment of loans on hand is already determined by their respective maturities, but anything that could be done to speed up or slow down the rate of making new loans would have power sooner or later to change the amount of loans on hand. For accomplishing this purpose one instrument which the banker has always at hand is the discount rate.

Reverting once more to the statement on page 553, let us suppose that the banker decides that his reserve ratio, thirteen per cent, is too small. He has been discounting, let us say, at six per cent, and for some time the new discounts have about offset those that were maturing. He therefore raises his discount rate to seven per cent. Since he is now charging a higher price for the service of discounting, fewer people will avail themselves of his services in accordance with the law of demand, and the making of new discounts will fall off. The old loans and discounts continuing to mature as before, the net result will be a decline in the bank's loans and discounts. Furthermore note that when customers pay off their notes they must pay either (1) in cash or in checks on other banks which will promptly be converted into cash, or (2) in the notes of this bank, or (3) in checks on their deposit accounts in this bank. Since loans and discounts are being paid faster than new ones are being made, the bank thus finds its reserve increasing and its demand liabilities decreasing. The reserve ratio thus rises. Let us suppose that after

a week's time loans and discounts have fallen off by \$100,000, cash has increased \$15,000, deposits declined \$80,000, and notes declined \$5,000. The result upon the bank statement, as compared with the condition on October 26, 1939 (see page 553), is as follows :

STATEMENT OF THE CITY BANK, NOVEMBER 2, 1939

<i>Assets</i>		<i>Liabilities</i>	
Loans and discounts	\$1,150,000	Capital stock	\$ 400,000
Securities	450,000	Surplus	75,000
Real estate	45,000	Undivided profits	12,327
Furniture	15,000	Notes	45,000
Reserve	207,410	Deposits	1,335,083
	<u>\$1,867,410</u>		<u>\$1,867,410</u>

The reserve ratio has been materially increased, being now $\frac{207,410}{1,380,083}$ or 15 per cent, as against 13 per cent. A continuation of the high discount rate will further tend to strengthen the bank's position.

Conversely a reduction of the bank's discount rate may stimulate business and cause new loans and discounts to come in relatively faster than the old are being paid off. Since the proceeds of the loans must be given in cash or in notes or in deposits, the tendency will be toward a decline in the reserve and an increase in the demand liabilities, thus lowering the reserve ratio. The rate of discount is thus, as it were, a control lever, by which the banker may regulate his business, steering a middle course between the desire for large profits and the danger of failure from inability to meet the demands of depositors and note holders.

Two qualifications should be noted in this connection. In the first place, while the bank has positive power to restrict loan expansion, its power to promote expansion is passive; it can make the borrowing terms more inviting to applicant borrowers, but the decision whether to borrow or not rests with the applicant. Oftentimes, especially in the later stages of business depression, low discount rates and other favoring acts of the banks do not avail to start revival. This principle was illustrated by the generally unsuccessful efforts to stimulate business by easy money and easy bank credit during the years from 1929 to 1938. In the second

place, control of the reserve ratio is not, in real fact, an important method in case of regular commercial and industrial borrowers at the banks. The banks use other methods of discouraging or encouraging applicants. When money is tight, commercial and industrial borrowers are able to get funds, if at all, at practically the same rates as when money is easy. Interest rates, except in highly organized markets, are very lethargic. The central bank rate and the rate for certain types of paper broadly dealt in do vary, for the purpose of control; but in ordinary times and in the ordinary bank the rate of discount tends to be customary and stable.

Note issue. The reader who has any practical knowledge of banking will not have failed to observe that in this discussion we are concerned primarily with the development of the fundamental principles of banking rather than with the description of any particular banking system. This distinction appears with particular sharpness in connection with the subject of note issue. This banking function has been so much regulated by legislation in the different countries, particularly in the United States, that the reader may have been surprised at the somewhat casual way in which we have thus far treated bank notes as though they were about the same thing as deposits. As a matter of fact the essential identity of bank notes and deposits is in the fundamental nature of banking; it is only legislation that has sharply distinguished them. This point needs some further elaboration.

Notes and deposits: Fundamental similarity. The bank deposit has already been defined as a right to receive money from a bank, evidenced by an entry on the bank's books and by other means. *A bank note is a bank's promise to pay money on demand.* The term "money" in both these definitions is of course exclusive of the bank's own notes. It will be noted that these definitions are almost identical. Deposits and notes are both forms of bank credit; both are liabilities of the bank; both are (except in the special case of time deposits) payable in money on demand. When a bank makes a loan it is immaterial to it whether the credit given the borrower is in the form of notes or deposits. To the borrower it is merely a question whether he shall receive the evidence of his right to receive money from the bank as an entry on the bank's ledger and in his

own pass book or inscribed on little pieces of paper which he will take away with him. These differences are superficial, not fundamental.

Differences between notes and deposits. There are however real differences between notes and deposits. The right to receive money from a bank, if in the form of a deposit, may be transferred only by check. Now a check is of itself no obligation of the bank. The person who receives it must be sure (1) that the signature upon the check is genuine and (2) that the maker actually has at least that amount on deposit and that it will still be on deposit when he presents the check to the bank for payment. Otherwise the check is not good. Checks are thus available for use only among people who are acquainted with one another or at least have confidence in one another. Their availability is also limited to those who are in the habit of doing business with banks. Checks are thus not generally convenient for payments between strangers, for paying traveling expenses, or for making ordinary wage payments. Where they are available, checks are usually a safer and more convenient means of payment than money. But after all they are not money, since they are not generally acceptable.

The bank note, on the other hand, is money. It is the definite obligation of a bank and, barring the risk of counterfeit or of the failure of the bank, may be accepted without any question as to the honesty of the one who tenders it. The bank note is generally acceptable; it may be used for all kinds of payments. Bank notes again are expressed in round sums, convenient for payments of all amounts. Although representing rights to receive money from the bank, bank notes may pass about for weeks or months or years before being presented to the bank for payment. They may travel far away from the bank of issue. The check, on the contrary, usually appears for the purpose of a particular payment, it is expressed in an odd amount of dollars and cents, and its life is ordinarily limited to that one transaction, after which it returns to the bank and is cancelled. That is, deposits require generally a new check for each separate transfer.

It results from these differences that checks and deposits are used by the commercial classes, bank notes by the laborers and poorer

classes who do not have deposit accounts. Most commercial transactions are settled by checks drawn against deposits, while notes are commonly used in payments between private individuals, and for small retail purchases, payment of hotel bills, purchase of railway tickets, etc. These distinctions between notes and deposits, though not so fundamental as their similarities, are yet real and important and justify the issue of bank credit in the two forms.

An important corollary of the differences between notes and deposits is that the opportunity for abuse and fraud on the part of an unsound or dishonest bank is far greater in the case of notes than of deposits. The deposits of a commercial bank are extremely active. Depositors are frequently demanding money, and deposits are being continually transferred by check, each check being promptly presented to the bank. The bank's condition is thus subject to constant test by persons having more or less direct contact with the bank and knowledge of its management and condition. The reverse is true in the case of notes, and an unsound or dishonest bank may long continue to defraud the public by means of its notes.

Legal restrictions. This is in part the explanation of the prevailing attitude of government toward the banking functions of deposit and note issue. In spite of their fundamental similarity and their original status as common law rights, the legislation of most countries has drawn a sharp distinction between them. While the banks have generally been permitted to handle their deposits with little interference by the government, the function of note issue has almost everywhere been subject to strict regulation. In most countries the right of note issue is restricted to certain kinds of banks or even to one particular bank. Thus in the United States only the federal reserve banks issue notes. In most of the leading nations of Europe, such as Great Britain, France, Germany, etc., the business of note issue is virtually the monopoly of a single great central bank, more or less closely associated with the government. Legislation further puts a limit to the amount of notes which the banks may issue, prescribes certain forms of security and a certain amount of cash reserve, determines the form and denomination of notes, etc. As a result the issue of bank notes is in most countries a

narrowly restricted and somewhat artificial function as compared with the business of deposit.

The problem of elasticity: Need of elastic currency. We have already observed ¹ that bank credit, in the Anglo-Saxon countries at least, performs the greater part of the medium of exchange function and so constitutes a very important element of the currency system. A peculiar and important function of bank credit is to furnish an element of elasticity to the currency system. No community's need of money is uniform at all times and seasons. In an agricultural community for example the volume of exchanges is especially great in the late summer and fall, on account of the business of harvesting and marketing the crops. Unusually large sums are being paid in wages to farm laborers, the farmers' receipts are at a maximum, and the local merchants find their sales swelling, while at the same time their credit customers are paying their bills. There is need of considerably more currency than during other seasons. In a manufacturing town the weekly pay envelopes present a special need of money on Saturdays. The payment of salaries on the first of the month and the prevailing habit of settling accounts at the stores and elsewhere during the first few days of each month make a monthly variation in the need of currency. Corporations very commonly make their payments of dividends and interest quarterly, usually on the first of January, April, July, October, which gives a quarterly swing to the fluctuating need of currency. There are also the cyclical fluctuations, the successive periods of business activity and business depression, of "good times and bad times." These examples should be sufficient to demonstrate that any community's need of currency is always a fluctuating quantity, the resultant of forces of the greatest variety and complexity. Adequately to meet this need, the volume of currency should fluctuate correspondingly and automatically.

It is true that the velocities of circulation of money and bank deposits tend to vary with the fluctuations of the business cycle and that therefore some of the variation in the cyclical demand for currency is met by responsive changes in velocities of circulation. This is not however sufficient to meet the situation. In the absence

¹ In Chapter XXII.

of an elastic element in the currency, variations in the volume of trade must be met by changes in the price level. Stability of prices is so important that there remains urgent need of an elastic volume of currency.

Elasticity of bank credit. Recalling the classification of money set forth in an earlier chapter,¹ and the essential characteristics of the several kinds of money, the reader will realize that no item in the standard money system, except bank notes, is so related to the state of business as automatically to expand and contract with the fluctuating need for currency.

With bank credit (notes and deposits) the situation is different. Recalling our discussion of the business of discount and deposit, it will be appreciated that the demand for loans at the bank arises directly out of business transactions, either recently completed or contemplated for the immediate future. When business is active the banks are called upon for many and large loans. And in accordance with the relation between loans and discounts on the one side and notes and deposits on the other side of the account, as the banks respond to the demand for loans they increase their issue of notes and their deposits. Currency thus expands to meet business needs. On the other hand, when business declines the customers of the banks have less need for currency. They pay off their obligations at the banks faster than new loans and discounts are made. In paying they either return to the banks the notes previously issued or they draw checks which reduce deposits. Currency contracts in response to declining need. Evidently there is here that close and automatic connection between volume of currency and business activity which is the mark of an elastic medium of exchange. Without such elastic element in its currency system, the modern community would suffer great inconvenience from alternating periods of scarcity and overabundance of currency. Bank credit thus performs a service of which the other parts of the currency system are generally incapable.

It is desirable that elasticity characterize both bank notes and deposits. Many people do not have bank accounts or do not understand the use of checks; in certain regions banks are few and scat-

¹ Chapter XXII.

tered and checks are awkward to use, and for small retail purchases bank notes are more handy. Inasmuch as the volume of all such transactions fluctuates, it is desirable that bank notes vary with the needs of the people. Unfortunately in the effort of certain governments to make bank notes safe their natural elasticity has been destroyed.

Influence of reserves. The elasticity of bank credit is limited by the necessity which is upon the banks to maintain adequate reserves. As has been seen, expansion of the banks' notes and deposits reduces their reserve ratio and inclines them to slow down the rate of expansion by raising the rate of discount. On the other hand, contraction of demand liabilities raises the reserve ratio, and the bankers seek to stimulate business by lowering the rate of discount. Banking reserves thus act, not to destroy the normal elasticity of bank credit, but to put on the brakes and so to prevent extremes of either expansion or contraction.

Other banking and credit agencies: The note broker. It sometimes happens that a business man chooses to make use of an intermediary in securing the funds he wants and prefers to place his paper in the hands of a *broker* for sale to some bank or institution seeking short-term investments. He may have discounted all the paper that his own bank will take and yet be in need of working capital, or he may find a broker who offers a lower commission rate than the discount rate of his bank.

One of the middlemen who has occupied a place of some prominence in the past is the *note broker*. He is a broker in the true sense of the word, in that his function is to bring buyer and seller together, and he receives a commission for so doing. The note broker still exists, but in recent years his place has largely been taken by the commercial paper houses.

Commercial paper houses. These are brokerage firms; they get a commission from the borrower for the services which they perform. They differ from the note broker in that they make loans directly to borrowers, whereas the note broker agrees to try to sell the commercial paper of his clients and in case he does not succeed in so doing returns the paper. But the commercial paper house has no intention of investing its own funds in the paper of its clients

save for the time which is necessarily spent in selling it. It expects to sell and as a general rule succeeds in selling to banks, insurance companies, or other companies and individuals that have funds available for short-term investment. The paper which it sells rarely runs for more than six months.

It may develop that the house has made a miscalculation of the absorbing capacity of the market, and it may be required to carry some of its clients' notes. In order to insure itself and also to protect the interests of those to whom it sells commercial paper it conducts a very rigid investigation into the affairs of each company. When it sells a note to a bank it does not guarantee its payment, but it does guarantee its genuineness and recommend its purchase. Large banking establishments usually conduct an independent investigation themselves, but the small country bank has to rely on the investigation of the commercial paper house or on the advice of its metropolitan banking correspondent.

Relatively little capital of its own is required by the commercial paper house. As has been stated, it buys this paper with the expectation of an early sale. Nevertheless each house must at any one time have a fairly large current investment in paper, and to carry this involves the temporary tying up of funds. These funds are secured by borrowing, at a lower rate of interest than that used to discount commercial paper of clients, from the large commercial banks on the promissory note of the house, secured by pledging the commercial paper of customers as collateral. Funds amounting to eighty or ninety per cent of the value of the collateral can be obtained in this way. If the house is forced by market conditions to carry some of the paper for a considerable time, it may make a profit by the difference in the rate of interest at which it borrows and the rate at which the customers' notes are discounted. There is of course the possibility of large losses, for the inability of the market to absorb the paper may be due to financial stringency, which will soon be evidenced by a rise in the interest rate, and to carry the paper the house may have to borrow money at a rate higher than the rate which was used in discounting it.

Such houses perform distinct services both to the bankers and to the business world. For the manufacturer or merchant a wider

and probably cheaper market is provided for his promissory notes than is offered by the local bank. He can get funds when the local sources of supply are exhausted. The bank, on the other hand, gains, in that it has a greater range of industries from which to select the paper it wishes to buy. Frequently a bank in a small city would otherwise be limited to the paper of one type of industry or to that of a group of middlemen handling the same product. Depression in that industry would cause more suffering to the bank than if it spread its risks (of non-payment) among different industries. It further enables it to select notes maturing at convenient times, making it possible therefore to change the nature of its investments from time to time.

In recent years there has been a relative and absolute decline in the volume of commercial paper. Note brokers and commercial paper houses are now handling more bank acceptances and less commercial paper, and the practice of the large corporations of accumulating large cash balances and of doing their financing through the issue of common stock rather than through commercial paper has decreased its volume. Since 1929 the need for borrowing through commercial paper houses has been relatively slight; industry and trade have been slack, and local rates of interest low.

Trust companies. Our attention has thus far been directed to the nature and functions of the commercial banks. This is by far the most important class of banking institutions, represented in the United States by the national banks and the state banks. Only a brief glance at the other kinds of banks will be required.

The trust companies are corporations whose original function was the care of their customers' investments and other financial affairs, the handling of trust funds, the administration of estates, etc. They naturally developed the business of keeping deposits subject to check and of making loans to their customers. Today the trust companies of the United States, in addition to their own peculiar functions, generally perform all the ordinary banking functions with the exception of note issue. To that extent the foregoing discussion applies to them equally with the strictly commercial banks. The differences between commercial banks and trust companies are steadily becoming less as time goes on.

Savings banks. These are institutions which accept deposits (usually in small amounts) and make investments for their depositors. They do not discount commercial paper, and their deposits are not the result of loans to their customers. Also their deposits are not legally payable on demand nor transferable by check in the ordinary way. The savings banks may require notice of withdrawal, of 30 or 60 days or some other period, though they frequently waive this right. Transfer of deposits requires certain formalities, such as the use of a special order and the presentation of the pass book. Savings deposits therefore do not serve as a means of making payments by check as in the case of the deposits of the commercial banks. The resources of the savings banks are invested, not in commercial paper, but in bonds, mortgages, and other investments of a safe and conservative character. The income thus obtained is devoted, after paying expenses, to the payment of interest to the depositors. The savings banks are really investment agencies rather than banks. Their customers are generally people in moderate circumstances, not engaged in active business. The amount that may be deposited by any one person is often limited, in order to restrict the savings banks' facilities to this class. For such persons the savings banks perform a useful service in gathering together small savings and investing them conservatively where they will yield a small but generally secure income. Wealthy persons and active business men have their own means of investing their funds and usually expect a higher rate of return. Recently there has been a movement among the commercial banks to enter the savings field, by establishing "savings departments." There is no corresponding tendency of the savings banks to take up commercial banking.

Private banks. Mere mention should here be made of the private banks, unincorporated banks conducted by individuals or partnerships and engaged primarily in the business of making investments, promoting corporations, underwriting security issues, etc. Some of them are to be counted among the most wealthy business concerns, and they occupy an especially significant position in international banking and finance, but they do not play an important part in the ordinary commercial banking world.

The function of credit: Credit in general. We may now profitably turn back to our starting point, in order to review the service rendered by credit, and particularly by bank credit. The importance of bank credit has been emphasized because it is fundamental in the economic organization, but it would be an error to ignore the service of commercial credit; that is, credit extended by the manufacturer to the wholesaler, or by the wholesaler to the retailer. In a great many transactions the banker is not called upon for credit. The manufacturer is able to take care of himself, or the producer from whom he gets his raw materials gives him time in which to work them up and pay for them.

We may in fact distinguish two points; namely, the importance of credit and the importance of a particular kind of credit. Credit, whether commercial or banking, is a sort of solvent factor in our organization, enabling us in a sense to make use of future goods at the present time. For society as a whole this cannot hold true of course, but it can for a given individual, for he gets present goods in return for a promise to pay in the future, and the transaction is conditioned on the belief of the lender in the borrower's ability to produce goods or services at some time in the future.

The gain to a given individual from the credit system may seem of little importance or interest, but there is a social gain as well as an individual gain. Recalling for a moment the situation illustrated in the example at the beginning of this chapter we can readily see that, if a system of payment on the delivery of the goods had prevailed at the time the consumers found themselves unable to pay for their goods, the retailer would have ordered no new goods from the wholesaler. The wholesaler would have ceased ordering goods from the manufacturer, who would have been forced to stop manufacturing. In a short time the consumers perhaps would have found themselves again in position to purchase goods on a cash basis, and renewed buying would have been passed on through the retailer and the wholesaler to the manufacturer. But how much more economical is that use of the productive resources which permits of a steady flow of goods instead of a jerky succession of activity and quiescence. Any business becomes somewhat disorganized if it has to shut down periodically, and disorganization is an expensive

thing for society. Credit, which enables some to make present use of the surpluses of others, facilitates that smooth working of the productive apparatus which is essential to economy. It is one of the factors which promote efficiency in production and which therefore tend to keep the output of consumable goods at a high level.

Peculiar importance of bank credit. Bank credit can perform a certain function in society better and more cheaply than other kinds of credit. The resources of any chain of producers are great, no doubt, but they are limited when compared with the resources of society as a whole, including the active leaders of business, those who are active in business as wage earners, and those who live on incomes derived from investments. The banks stand at the central point of the financial world and serve as reservoirs which attract the savings of all. They are in touch with more people and more lines of industry than the typical firm can possibly be. If there is for the time being a surplus of funds in one line of industry, it can be made available for other industries. Without the bank as an intermediary, how for example would the proprietor of a cotton mill in Massachusetts know that the owners of brass factories in Connecticut had funds which would be idle for six months? And if he did discover this fact, how could such funds be made available to him? It might be done, but it would be difficult and costly. The bank stands ready to receive the funds of the Connecticut manufacturer and to pay him interest on a time deposit for six months; it can loan these directly to the cotton manufacturer in return for a promissory note of six months' duration. The thing is accomplished directly, simply, and at little expense.

The student will realize of course that the banks do not act only as distributors of funds which are deposited with them. Actually their function as creators of credit is more significant and more important quantitatively. When a bank gives a borrower bank notes or a deposit, it increases the quantity of purchasing power in circulation and thus enables him to secure present goods or services on the basis of his promise to pay.

Furthermore the bank serves as a selective factor in the direction of funds. If funds are needed more urgently in one industry than in another they are directed toward that industry, for the greater the

need the higher the rate of interest which will be paid to secure a loan, and, other things being equal, a bank in making its loans will give preference to those willing to pay the higher rate of interest.

EXERCISES

1. Construct a bank statement from the following items: Real estate, \$85,000; Capital stock, \$100,000; Stocks and bonds, \$44,000; Surplus, \$45,000; Undivided profits, \$20,000; Loans and discounts, \$420,000; Deposits, \$480,000; Reserve, \$96,000. What is the bank's reserve ratio?

2. The bank discounts a note for \$10,000 for 3 months, the rate of discount being 6 per cent per annum. The proceeds are credited to the borrower's deposit account. Make the necessary changes in the bank's statement. What is the reserve ratio now? *NOTE:* In this and each of the following four exercises, refer to the original statement as in Exercise 1.

3. The bank receives deposits of \$3,000 in cash and \$5,000 in checks drawn upon other banks. Make the necessary changes in the bank's statement. What is the reserve ratio now?

4. The bank purchases securities in the amount of \$10,000. Make the necessary changes in the bank's statement. What is the reserve ratio now?

5. A customer repays a loan of \$8,000 by check drawn against his deposit account in this bank. Make the necessary changes in the bank's statement. What is the reserve ratio now?

6. The bank rediscounts a note for \$10,000, 2 months before maturity, at the federal reserve bank. The rediscount rate of the latter bank is 3 per cent, and the proceeds are taken in notes of the federal reserve bank. Make the necessary changes in the bank's statement. What is the reserve ratio now?

7. The clearing house association in a certain city includes five banks. On a given day they present checks at the clearing house as follows:

<i>Checks presented by:</i>		<i>Drawn against:</i>			
	Bank A	Bank B	Bank C	Bank D	Bank E
Bank A	—	\$2,100	\$1,800	\$2,400	\$1,600
B	\$1,900	—	1,500	2,100	1,800
C	2,300	1,600	—	1,800	1,500
D	2,600	1,800	2,000	—	1,700
E	1,400	2,000	1,400	1,900	—

Determine the balance to be paid or received by each bank.

XXV

THE AMERICAN BANKING SYSTEM

Historical background. The banking system of the United States is the result of continuous experimentation, by private individuals and the governments of colonies, states, and the nation, in furnishing money and credit to the people; it bears the earmarks of struggle and compromise among the powerful forces that have featured American national development. Chief among these forces have been insistence upon democratic ideals and local self-government, with corresponding reluctance to concede authority to the federal government, and the fear lest control of the institutions of money and credit be monopolized in the hands of a few. The general result emerging from this background is a loose decentralized system of thousands of small local unit banks, each being chartered and supervised by one of the forty-eight states or the federal government. At the head of the system are twelve central banks, federated under a scheme of shifting authority and responsibility though at best representing less than half of the unit banks. The skeptic is wont to question the propriety of calling such a conglomeration a "system," a term connoting order, articulation of parts, and central control. A true understanding of this "system" can best be acquired by the aid of a study of its background.

Banking in the United States before the Civil War. The so-called "banks" of the American colonial period were not real banks, but merely "batches of paper money" as one historian puts it; they did not engage in the business of discount and deposit and even in the issue of their notes did not conform to the principles of banking which our study has developed. But during the Revolution and the years immediately following a few real banks were founded, which in addition to issuing notes made a start in the field of discount and deposit.

In 1791, two years after the formation of the national government, Congress established a central bank, the first Bank of the

United States. This bank performed all the normal functions of a central bank, was the fiscal agent of the United States government, and did a regular banking business with the general public in competition with the growing company of local banks. Its notes became the principal circulating medium of the country. It was wisely and conservatively managed, performed its various functions well, and exercised also a salutary influence over the local banks, whose tendency to wild and unsound methods was apparent from the first. Indeed it was this check upon the ambitions of the local bankers and its competition with them that finally proved the bank's undoing. Opposition arose, and Congress eventually failed to renew its twenty-year charter, which expired by limitation in 1811.

For the next five years the nation's banking was wholly in the hands of the local banks, chartered by the several states. These banks increased prodigiously in number. Relieved of the restraining influence of the great central bank, the majority ran wild in the paths of reckless and unsound banking. The country was flooded with their notes, mostly of doubtful redeemability, they failed to perform satisfactorily the fiscal services which the Bank of the United States had regularly provided the government, and finally in the financial panic of 1814 the whole banking structure went to pieces, a condition made doubly disastrous as it came in the midst of the War of 1812.

The people learned the lesson, their misgivings against a strong central bank were temporarily overcome, and in 1816 the second Bank of the United States was founded. This bank's record is in its essential features a repetition of that of the first Bank of the United States, including its notable services to the government and the community and its final engulfment in the bog of politics. Its charter was not renewed, and it closed its doors in 1836. There was now a repetition of the experiences of 1811-1816, and the banking system collapsed once more in the panic of 1837.

Reestablishment of a central bank proving politically impossible, the United States government cut itself loose from banking altogether by the establishment (first in 1844 and finally in 1846) of the Independent Treasury for the custody of all government funds

and the management of the government's financial business. The public was left to rely for banking services upon the local banks. Gradually some order came out of the chaos, and in the older and more settled parts of the country, particularly New England, New York, and Indiana, many sound banks arose and became more or less closely associated in an efficient banking system. "Wildcat" banking continued to prevail in the frontier regions, and the country suffered grievously, especially from the multiplicity of note issues of all degrees of goodness and badness. Thus things continued down to the time of the Civil War.

The national banking system: Origin. The Civil War ushered in a new chapter in American banking history. Laws passed in 1863 and 1864 provided for a new class of banks chartered by the United States government, thus inaugurating the *national banking system*. The purpose of Congress in this legislation was primarily to aid the financing of the war and only secondarily to reform the banking system. In harmony with the first purpose, the law provided (as amended after the Civil War) that each of the new national banks with a capital in excess of \$150,000 must, before commencing business, purchase and deposit with the Treasurer of the United States bonds of the United States to the amount of at least \$50,000 and that each national bank with a capital of \$150,000 or less must deposit an amount of bonds equal to at least one fourth of its capital. Upon the security of the United States bonds so deposited the bank was permitted to issue its notes in any amount not in excess of the par value¹ of the bonds or the capital of the bank. National banks also issued notes secured by the deposit of "lawful money."

By an act of 1865 Congress imposed an annual tax of ten per cent upon the notes issued by state banks, thus making note issue unprofitable and reserving that function to the national banks. There was also relaxation of the Independent Treasury system to the extent of permitting deposit of government funds in selected national banks and the receipt and disbursement of national bank notes by the Treasury.

¹ Originally and up to 1900 the limit was 90 per cent of the par value or the market value, whichever was smaller.

After the first few years the organization of the new national banks proceeded rapidly, and they soon assumed a position of importance in the American banking system, which they have since maintained. The national banks, the state-chartered banking institutions (state banks, trust companies, and saving banks), and the private banks have existed side by side, dividing among themselves the country's banking business.

National bank notes. It will be observed that the notes of the national banks were not issued freely upon the basis of their general assets, in accordance with the simple banking principle discussed in the preceding chapter, but were secured by a particular class of assets set aside for the purpose with a trustee. The United States law further provided for the safety of the national bank notes by imposing upon the United States Treasury the obligation to redeem them on demand, with of course the Treasury's consequent right to collect the amount from the bank of issue. For this purpose each bank was required to maintain a cash fund with the United States Treasury equal to five per cent of its outstanding notes. The holder of a national bank note therefore had no concern as to the soundness of the bank of issue, relying as he did upon his claim against the United States, and the deposit of bonds was primarily to safeguard the Treasury rather than the note holder.

The national banking act was not very successful, during the Civil War, in accomplishing its main purpose of increasing the demand for United States bonds and raising their value. Later on however it did exactly that. The national banking system expanded, and the national debt was reduced, so that by 1911 the banks were using three-fourths of the entire interest-bearing debt of the United States to secure their note issues. In that year, the defects of the bond-secured notes having become evident, the United States adopted the policy of making its future loans by means of bonds and notes not available for bank note security, and the national banks were limited to those classes of bonds only which were in existence on May 31, 1911. Here there was an artificial market, which regularly absorbed nearly the whole stock of the United States bonds eligible for securing bank notes and gave to them a value far above their worth as an investment to others

than national banks. On June 30, 1928, for example, the national banks were using as security for their notes 98.67 per cent of all the United States bonds that were eligible for that purpose. All eligible bonds appearing on the market were snapped up by the banks, and their prices were driven above par.

While the device of bond security made the national bank notes eminently safe, this was at the expense of elasticity. Since government bonds had to be bought and deposited before additional notes might be issued, the banks could not quickly expand their note issues in time of need. On the other hand, the artificially valued bonds could not be sold in large volume to purchasers except other national banks save at a serious discount, and the banks were for the most part unwilling to hold such low-rate bonds without having the full amount of notes outstanding. One can readily see that there was no likelihood of large contraction or expansion of the volume of these notes consonant with the activity of business. On the other hand, the reader will not fail to recall that during the last quarter of the nineteenth century the nation was passing from note currency to deposit currency and that the importance of an elastic system of bank notes was gradually decreasing.

The problem of reserves. Because of the great extent of territory, the wide diversity of industries, and the predominant position of agriculture, the need of an elastic currency and of centralized reserves is particularly urgent in the United States. Until the foundation of the federal reserve system in 1913, this need was not met. The United States has been *par excellence* the country of small independent banks; by 1913 the nation's banking business was in the hands of some 26,000 local banking institutions, with an average capital of about \$80,000 and average resources of about \$775,000. Branch banking was until recently forbidden by law to the national banks and in most states to the state banks and trust companies. In the absence of any system of centralized banking reserves, such as has been provided in most countries by the central bank or by a small group of large and powerful banks with their far-reaching system of branches, each bank was forced to rely chiefly upon its own resources to meet whatever situation might confront it.

The national banking act required each bank to maintain a legal reserve against deposits, but for all banks except those in certain large cities a portion of this reserve might be in the form of deposits with national banks in certain of the larger cities. Quite naturally the banks in the country districts and smaller towns and cities developed the custom of depositing as much of their resources as they felt they could spare in periods of slack business with banks in the large cities of their respective regions. These city banks in their turn made deposits in other larger cities. Thus were set up, as it were, little rivulets of currency springing from sources all over the country, uniting in the cities to form larger streams, joining again into still larger, till eventually a mighty flow of money poured into the banks of a few financial centres, the greatest of all being New York City. This flow of deposits was encouraged by the willingness of the large city banks to pay interest at a moderate rate upon deposits of other banks. The funds so obtained were loaned by the city banks to their customers, a considerable part being used for stock market speculation. On the other hand, when a succeeding period of business activity opened up larger opportunities for the outlying banks, or when a crisis threatened, they demanded the return of their reserves, and the flow of currency was reversed. This forced the city banks to call in their loans and contract their deposits and caused a state of monetary stringency in these cities. This ebb and flow, besides its minor variations, had a great annual cycle on account of the need of currency for harvesting and marketing the crops. In the spring there was a period of easy money, a flow of currency to New York and a few other financial centres, low interest rates, and encouragement to speculation in these cities. In the late summer and fall money flowed back to the country, while in the large cities monetary stringency prevailed and interest rates soared.¹

It will be seen that, in a sense, the great city banks, particularly in New York, undertook to perform the services of a central reserve reservoir. But they were fatally handicapped by their great

¹ Statistical studies have shown that there was no dependability in the periodicity, either by season or business cycle, and the correlation of interest rates, stock prices, and money movement. Had there been, it would probably have been largely eliminated by speculation.

number, their independence of each other, their lack of any branch organization, and their inability to expand and contract their credit in the form of notes. Moreover they were unwilling to accept the responsibility of their position to the extent of keeping adequate reserves. The result was inevitable. Alternating periods of surplus currency and monetary stringency were the normal thing; inability of the outlying banks to obtain the return of their reserves in time to avoid disaster was common; bank failures and business failures resulted; financial panics and crises were made more frequent and more severe. The situation grew steadily more intolerable with the development of the country's wealth and industry, until at last a remedy was sought in the federal reserve system, which became law on December 23, 1913.

The Federal Reserve: Organization; Board of Governors. The federal reserve system does not supplant the national banking system but is superimposed upon it. It was intended to furnish the United States with a centralized banking organization, performing in general the functions of a central bank, and to give to the banking system as a whole a unity and a coherence which it previously lacked.

At the head of the Federal Reserve stands the Board of Governors of the Federal Reserve System (known until 1935 as the Federal Reserve Board) with its headquarters in Washington. At present the Board consists of seven members appointed by the President with the advice and consent of the Senate. Their salaries are \$15,000 a year and they are appointed for terms of fourteen years. The appointments are so staggered that a vacancy occurs every two years. In his selection of members of the Board the President is required to give due regard to the fair representation of financial, agricultural, industrial, and commercial interests. To ensure the various geographical districts of the country representation on the board, the law provides that not more than one member may be appointed from any given federal reserve district. The President designates one of the members to serve as chairman of the Board of Governors for a term of four years.

Federal reserve districts and banks. The whole country is divided into twelve "federal reserve districts," with a "federal

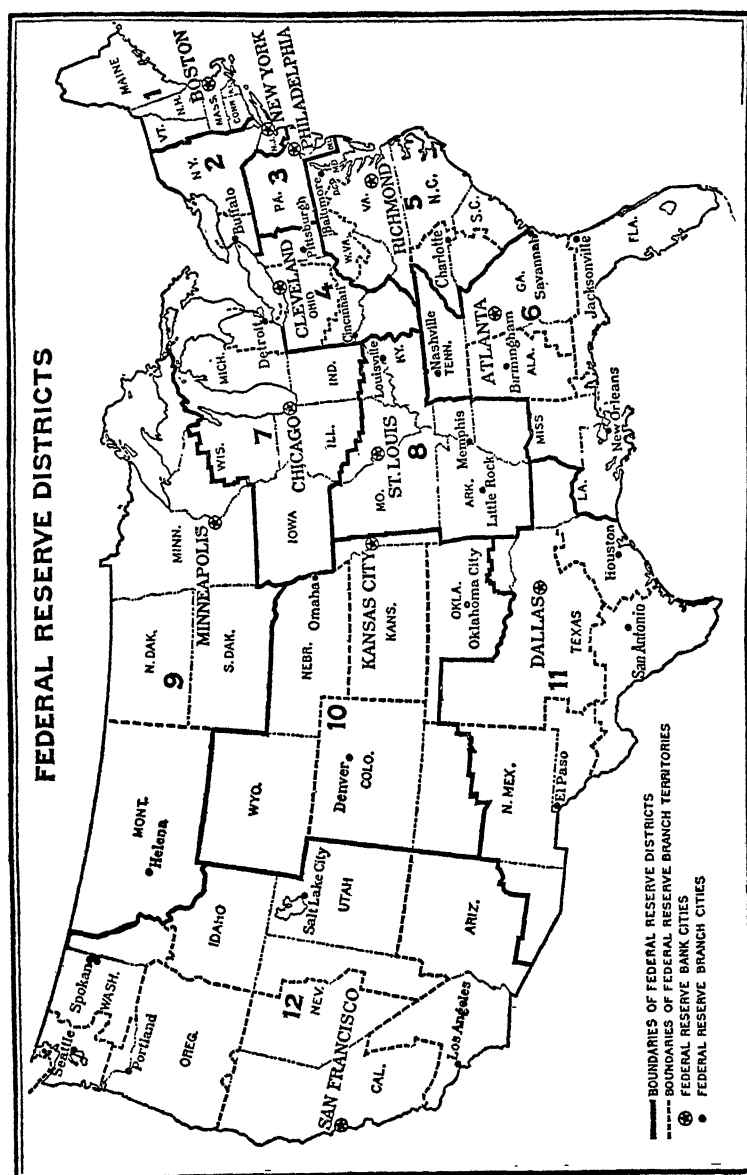


Fig. 36

reserve bank" located in a certain city in each district. The districts are distinguished by numbers, the following being a list of the districts and their respective federal reserve cities: 1, Boston; 2, New York; 3, Philadelphia; 4, Cleveland; 5, Richmond; 6, Atlanta; 7, Chicago; 8, St. Louis; 9, Minneapolis; 10, Kansas City; 11, Dallas; 12, San Francisco. Each federal reserve bank is permitted to establish branches and agencies located in cities in its district and abroad. At present there are 25 branches. The accompanying map (Figure 36) shows the boundaries of the several districts and the location of the federal reserve banks and their branches.

A board of nine directors is entrusted with the management of each federal reserve bank. Six of the directors are chosen by the member banks, three to represent the interests of the member banks and three the interests of the business and industry of the district, while the remaining three directors are appointed by the Board of Governors. One of the latter is designated "federal reserve agent" by the board and serves also as chairman of the board of directors of the bank.

Member banks. Each federal reserve bank is a corporation, and its stock is all owned by the individual banks within its district which have come into the system. These banks are known as "member banks." Every national bank is required by law to be a member, but membership is optional for the banks and trust companies incorporated under the laws of the states. On September 28, 1938, there were 6,341 member banks, of which 5,239 were national banks and 1,102 were state banks and trust companies. Less than twelve per cent of the state banks and trust companies were members, but these represented 40 per cent of the total resources of all the state banks and trust companies, showing that the largest and strongest institutions have pretty generally come into the system.

Functions of the Federal Reserve: Rediscounting and discounting. The banking business of the federal reserve system is conducted by the twelve federal reserve banks. They rediscount paper for their member banks and for each other, they hold deposits of the member banks, they issue notes, and they act to a certain extent as fiscal agents of the United States government. They are

permitted also certain dealings with the general public, but their principal business is with their respective member banks, with each other, and with the government.

Rediscounting by the federal reserve banks is in harmony with the general principles with which the reader is familiar. The law makes certain restrictions as to the kind of paper that is eligible for rediscount, prescribing, up to 1935,¹ that it must be of short duration and based upon commercial, industrial, or agricultural transactions or upon loans secured by United States government obligations as collateral and bearing the endorsement of a member bank. A member bank rediscounting paper with the federal reserve bank of its district may receive the proceeds either in the form of a deposit with the bank or in money — usually federal reserve notes.

A member bank may also procure federal reserve notes or other money from the federal reserve bank or increase its deposit account in the bank by getting so-called “advances,” that is, by discounting its own note backed by government securities or eligible paper as collateral.

Federal reserve notes. Two kinds of notes may be issued by the federal reserve banks, federal reserve notes and federal reserve bank notes. The federal reserve note, though nominally the obligation of the United States government, is the direct obligation of the federal reserve bank of issue, constituting together with the federal reserve bank notes a first lien against all the assets of the issuing bank. These notes are thus essentially bank notes of the normal type described in the preceding chapter, rather than government notes. Under the provisions of the original act they could be issued only on the basis and to the amount of eligible paper (as defined above ²) pledged with a trustee — the federal reserve agent — to secure them. In addition the bank of issue was required to hold a gold reserve, which must be not less than 40 per cent of the amount of the notes outstanding, though there was provision for going below this minimum in time of emergency subject to a progressive tax on the deficiency.

¹ The important changes made in this feature by the banking act of 1935 are described in the next chapter.

² Paper based on government securities as collateral did not come within the original definition of paper eligible for rediscount.

As provided in the original act, federal reserve notes go into circulation when member banks rediscount their customers' paper at the reserve bank, which in turn procures notes by pledging the discounted items with the federal reserve agent. Notes normally go out of circulation when customers repay their indebtedness, for this enables the member bank to liquidate its borrowings at the reserve bank and forces the reserve bank to retire an amount of notes equal to the repaid loans, unless indeed it has other paper to pledge with the reserve agent for that which has matured and been paid off. Another provision of the law, which forbids a federal reserve bank to pay out the federal reserve notes of any other reserve bank under penalty of fine, provides the mechanism by which the notes of each federal reserve bank are being continually returned to it for retirement or reissue.

Elasticity of the federal reserve notes. The founders of the federal reserve system contemplated that the federal reserve notes would ultimately constitute the largest part of the monetary circulation, and they believed that elasticity of note issue, with the volume of notes responsive alone to the seasonal, cyclical, and other changes in business needs, was assured by the above described relation between the notes and eligible commercial paper. As matters have developed, while the federal reserve notes are as safe and sound as the credit of the government and the assets of the federal reserve banks can make them, the scheme of elasticity has been seriously upset by a number of amendments to the federal reserve act.

In order to assist the government in placing loans, the federal reserve banks were in 1916 permitted to make advances to member banks on their obligations secured by government bonds, thus doing away with the assurance that such borrowings were altogether for commercial purposes. In 1917 the direct issue of notes in exchange for gold was permitted, which meant that thereafter the volume of notes was affected by the production and net importation of gold. Furthermore the requirement of 140 per cent backing for the notes was altered. Notes might henceforth be secured by 60 per cent commercial paper plus 40 per cent gold, by 100 per cent gold, or by any other combination of paper and gold between these

two. Finally in 1935 the reserve banks were authorized to regard government bonds as eligible paper which might be pledged with the federal reserve agent for federal reserve notes. This last amendment expires in 1939, but for the time being it has displaced completely the founders' plan for elasticity.

Since the sequestration of gold, described in Chapter XXIII, gold no longer serves as a part of the reserve but is replaced by gold certificates on hand or due from the United States Treasury. On January 25, 1939, the federal reserve agents held as collateral against an issue of 4,686 million dollars of federal reserve notes the following items: 4,791 millions of gold certificates on hand or due from the Treasury and about 4 millions of eligible paper. In other words practically the entire collateral held against outstanding federal reserve notes was in the form of gold certificates.

While the founders' plan for elasticity has thus gone askew, it must not be thought that the volume of federal reserve notes in circulation is invariable or constant. The volume is affected by the need for hand-to-hand money, which varies with business activity, normally reaching a peak during the Christmas holidays. Any member bank may always meet an additional need for currency by translating its deposits with the reserve bank into federal reserve notes. Moreover by the process of rediscount or borrowing on its collateral note, the member bank in sound condition may always increase its deposit with the reserve bank or its stock of federal reserve notes. The local banks can meet the needs of their customers for currency in times of special business activity through granting deposits or giving federal reserve notes; when the special need is over the currency contracts as loans are paid off, deposits decline, and the federal reserve notes are returned to the federal reserve banks in exchange for paper previously rediscounted and now paid off.

Plan to retire the national bank notes. It was evidently the intent of the authors of the federal reserve act that the notes of the federal reserve banks should ultimately displace at least the greater part of the existing national bank notes. The law therefore repealed that part of the national banking act which required each national bank to invest a certain part of its capital in United

States bonds. It further provided that, within certain limits, the national banks might sell to the federal reserve banks at par bonds which they had deposited to secure their notes and might retire the corresponding notes. The federal reserve banks, upon purchase of such bonds, were permitted to issue their own notes secured by the bonds. These notes are called *federal reserve bank notes* and are in all essential respects like the national bank notes which they replaced. As an alternative to issuing federal reserve bank notes, the federal reserve banks were permitted to exchange the two per cent bonds, thus bought from the banks, with the United States Treasury for new three per cent bonds or notes without the circulation privilege. As a matter of fact this program was never carried out; the interference of the World War and the later reluctance of the national banks to surrender their circulation made it abortive.

The federal reserve bank notes have never become an important part of the monetary system. On May 31, 1938, there were only 31 millions outstanding, and these were a residue from an emergency issue during the depression and had nothing to do with the process of retiring national bank notes. Moreover since March, 1935, the obligation of the federal reserve banks for their outstanding bank notes has been extinguished by the deposit of funds with the Treasury, and as the notes are turned in they are being retired.

End of the national bank notes. The fate of the national bank notes was determined otherwise than according to the 1913 plan. On March 11, 1935, the Secretary of the Treasury gave notice that all outstanding Consols of 1930 and Panama Canal bonds of 1936 and 1938 were called for redemption on August 1, 1936, using for this purpose the greater part of the gold "profit" from devaluing the gold dollar which was not devoted to the "stabilization fund." The entire issue of two per cent bonds, which for many years had been the only bonds to confer the circulation privilege, was thus wiped out by this action. The right to issue national bank notes against certain other United States obligations, which had been conferred by legislation of July 22, 1932, expired on July 22, 1935. In consequence there existed no bonds eligible for securing national bank notes after August 1, 1935. The Treasury has paid the

national banks for their bonds and returned to them their redemption fund, and the banks have paid to the Treasury the sums necessary to redeem their notes. As fast as the national bank notes come into the Treasury they are redeemed and cancelled. Thus closes one chapter in American monetary history. This action results in some simplification of the currency system and reduces somewhat the advantage of the national bank charter.

Reserves against deposits. The federal reserve banks receive ordinary deposits only from their member banks, foreign banks, and the United States government and certain of its agencies, but deposits for the purpose of exchange and collection are received also from other federal reserve banks and from state banks and trust companies which are not members of the system. Against its deposits each federal reserve bank is required normally to keep a reserve of gold certificates (formerly gold) or lawful money of at least 35 per cent.

Each member bank is required to keep with the reserve bank of its district a deposit not less than a certain percentage of its own deposits; namely, 5 per cent for time deposits and $22\frac{3}{4}$ or $17\frac{1}{2}$ or 12 per cent for demand deposits, depending on whether the member bank is located in a "central reserve city," a "reserve city," or elsewhere. This is called the member bank's *lawful reserve*. The Board of Governors has power to alter these reserve ratios, provided that they may never be set lower than 13, 10, or 7 per cent or higher than double these percentages. All moneys in possession of the member bank, as well as its demand balances with other banks, serve of course as working reserve, but such funds are not part of its lawful reserve as technically defined.

Open-market operations. The dealings of the federal reserve banks with the public, called *open-market operations*, consist of buying and selling certain specified items — bankers' acceptances, government securities, and gold — in the public markets.¹

¹ As a result of their rediscount and open-market operations the twelve federal reserve banks held on November 20, 1929, taking this date for illustration, the following:

Bills discounted, secured by government obligations	\$429,160,000
Other bills discounted	470,398,000
Total bills discounted	\$899,558,000
Bills bought in open market	283,831,000

If a federal reserve bank buys government securities or eligible paper from a member bank, the effect of the operation is to relieve tension in the money market. If the member bank's reserve balance is credited with the proceeds of the sale, it is quite evident that the bank's lawful reserve is increased and it is in a position to lend further. If the member bank instead takes the proceeds in the form of federal reserve notes or other money, its vault cash is increased and it may be disposed to lend more freely thereafter.

If, on the other hand, the federal reserve bank buys the government securities or eligible paper from a non-member bank, the effect of the operation is quite similar. The federal reserve bank may give the non-member bank a check on itself for the proceeds of the sale. The non-member bank will send the check to its metropolitan correspondent for collection and credit to its balance with the correspondent, and inasmuch as this balance is regarded in law and in fact as legal reserve, the non-member is now in position to extend loans more freely.

Finally if the federal reserve bank buys the securities or paper from a private person or non-banking corporation, the check received by the seller will be deposited in his bank (member or non-member) and, when collected, will increase the local bank's reserve by the amount of the proceeds of the sale.

The sale of securities by a federal reserve bank, whether paid for by federal reserve notes or other money, or by a check on the federal reserve bank, or by a check on a member or non-member bank, will tend to restrict the amount of money or credit in circulation. Money may be actually withdrawn from circulation or member banks' reserve balances with the federal reserve banks will be decreased, with the result of reducing the power of the banks to grant loans to their customers.

Open-market operations, together with credit rationing and moral suasion, have come to be relied upon by the federal reserve banks more largely than changes in the discount rate as methods of controlling credit, in marked contrast with the Bank of England, the Bank of France, and the other great central banks, whose influence over the national discount market is exercised largely through manipulation of their own discount rates.

Centralized reserves. It will be apparent from the previous account of the business of the federal reserve banks that the system is fairly well adapted to furnish a central reservoir and centralized control for the country's banking reserves. The only reserves against deposits now required by law of the member banks are their balances with the federal reserve banks. Of course the individual banks must keep some cash reserves on hand, in order to meet their daily obligations, but this is left to the discretion of the particular bank, and the amount actually kept is extremely small. For example, on September 28, 1938, the 6,341 member banks had together 775 million dollars "cash in vault" against "demand deposits" of 30,309 millions and total deposits of 41,776 millions. This was 2.5 per cent of demand deposits and 1.8 per cent of total deposits. Their "lawful reserves" on deposit with the federal reserve banks on this date were 8,195 millions, 27.0 per cent of their demand deposits and 19.6 per cent of total deposits. The banks also find it expedient and even necessary to carry balances with correspondents. During the latter part of the depression of 1929-36 and again in 1938 the reserve situation was very abnormal, inasmuch as the member banks carried heavy excess reserves — on the above date 2,919 millions — whereas in normal times they are close to the legal minimum plus necessary till money.

The federal reserve banks have, on the other hand, maintained very large reserves, generally far in excess of the legal requirements. Their gold holdings were almost from the start extraordinarily swelled as a result of the suspension of gold payments throughout most of the world, the enormous demand for American exports, and other events of the World War. In the wake of the War the favorable trade balance remained large; moreover the political instability of Europe and other parts of the world, together with the inflationary programs of public finance employed in certain countries, inclined many people to try to preserve their fortunes by shifting the fluid part to America. The great boom in the American securities market in the late 1920's accentuated this tide of funds. The flow of gold to America was a major factor in breaking down the gold standard in Europe in 1931, when Eng-

land, followed by most of the other countries, resorted to schemes of irredeemable currency.

The successive steps by which the United States in 1933 and 1934 departed from the gold standard, devalued the dollar, and expropriated all the gold in the nation have been described in an earlier chapter.¹ The federal reserve banks were given, in exchange for the gold taken from them, a special form of gold certificates or a credit at the Treasury. Since that date their gold holdings — the gold reserves of the federal reserve banks against their notes and deposits — have consisted of “gold certificates and due from the U. S. Treasury.” Many other countries also devalued their currencies, set up barriers to international trade and finance, and embarked upon programs of extreme nationalism. These policies, along with political instability and wars and threats of war, have stimulated the flow of huge amounts of gold to the United States since 1933.

The record of the gold reserve of the federal banks, in the light of this series of events affecting it, is shown in the following table :

<i>Date</i>	<i>Gold reserves (millions of dollars)</i>	<i>Ratio of gold re- serves to combined amount of note and deposit liability</i>
April 27, 1917	522	62%
May 28, 1920	1,969	35%
December 31, 1933	3,524	59%
June 22, 1938	10,636	80%

Mobile reserves. Not only do the federal reserve banks have a large reserve, in absolute amount as well as in relation to their liabilities and to the total monetary gold of the world, but this reserve, as well as the reserves of the member banks, is effectively mobilized. The banking reserves of the country are concentrated, available, and controlled in the public interest. Through dealings between the reserve banks themselves, directed and controlled by the Board of Governors, it is possible to direct the flow of reserves to the particular districts where for the time being the need is urgent. Whereas in former days sound liquid banks were sometimes forced to fail because they could not obtain quickly the money

¹ Chapter XXIII.

due them from other banks, such an event is now virtually impossible. A local panic can no longer wreck a sound liquid member bank or a series of banks by a wild unreasoning "run" of depositors to withdraw their money. Any soundly managed member bank can convert resources into cash sufficient to meet any run, either by rediscounting with the federal reserve bank certain eligible paper — and the definition of eligible paper was very much widened by the banking act of 1935 — or by borrowing on its notes with United States government securities as collateral. On the other hand, a bank which has no eligible paper or government securities — a condition clearly indicating unsound management — has no means of approach to the federal reserve bank.

State banks that are not members of the federal reserve system cannot, except by special dispensation of the Board of Governors in emergencies, get help directly from the reserve banks; they may however be able to borrow from a correspondent bank which is a member of the federal reserve system, and this member may procure loan funds from the reserve system for the purpose.

A national system of clearing. The federal reserve system has also become a national clearing house for settling the claims of the reserve banks against each other and for collection of checks, drafts, and other claims of individual banks against out of town banks both in its own and in the other federal reserve districts.

The federal reserve collection system is however far from universal; in fact it has declined in universality since 1922, when maximum efforts were made to extend it. For instance on December 31, 1938, 11,973 banks, 81 per cent of the total, were on the par list. A bank on the par list is one that will remit at par for checks sent to it by the federal reserve banks for payment. All member banks are legally obliged to do this; the other banks may voluntarily agree to remit at par, or the federal reserve bank may find means through another bank in the town to collect at par from them. Nevertheless there were on this date 2,760 banks from which the federal reserve banks were still not able to collect at par.

When checks are sent to the federal reserve bank for collection, credit is made to the depositing bank's collection account at par,

but the amount of the check does not become available for draft or for reserve purposes until time has elapsed sufficient on the average to allow for its collection and the remittance of the proceeds to the federal reserve bank. Each bank which belongs to the federal reserve collection system may collect through the federal reserve banks to the degree it wishes. The system would attain the maximum of efficiency if all banks belonged to the system and used it exclusively ; but as it is at present certain small banks stay outside the system, in order that they may continue to make exchange and collection charges and thus add to their earnings. In spite of its imperfections however this mechanism is far more effective than anything that ever existed in the country before. It secures collection generally at par without charge in the shortest possible time and with a remarkably small transfer of actual money, thus doing away with what was formerly a cause of much delay and heavy expense to the public.

The government's financial business. The reader will recall that, following the banking collapse in 1837, Congress established the Independent Treasury system, whereby the United States government undertook to keep its own moneys and handle all its financial business without the aid of the banks. It will also be recalled that with the later establishment of the national banking system the Independent Treasury law was relaxed to the extent of permitting deposit of government funds in selected national banks. The government thereafter conducted its financial business through the Treasury at Washington, sub-treasuries located in nine principal cities, and the offices of the mint, with the aid of a variable number of national bank depositories. The system was always an anomaly, unlike anything to be found among other important nations, and fraught with danger to the country's monetary stability. Use of national bank depositories, chosen from the thousands of such banks, was inconvenient, wasteful, and politically insidious.

The federal reserve act partly removed the distinction in favor of national banks by extending the privilege of holding government deposits to state banks that were members of the Federal Reserve. Of greater significance was the provision that the funds of the

United States (with certain exceptions) might be deposited with the federal reserve banks and that these banks might be required to act as fiscal agents for the Treasury. The Treasury was at first slow to avail itself of these facilities, but the exigencies of the World War, in particular the conduct of the government's loans and the handling of the funds so obtained, soon compelled the utmost reliance upon the Federal Reserve, whose chief activities became concentrated upon its function as fiscal agent of the United States. At the same time the number of local bank depositories was greatly increased, and many "special depositories" are still used in connection with the government's borrowing operations. In 1921, under authority of a law of 1920, the sub-treasuries were abolished as no longer required, and the Treasury made progress, slowly to be sure, toward curtailing its reliance upon the local banks and concentrating its fiscal business in the hands of the federal reserve banks.

The American banking system in 1929. The federal reserve banks have proved themselves an important addition to the American banking system. They assisted in the difficult task of financing the country during the World War, and they have given the banking system more semblance of order and centralized control than it ever had before, but there still remain serious defects, both in the structure and in the functioning of the American banking system, which the Federal Reserve has failed to remedy. Even in the years of prosperity prior to the crisis of 1929 the existence of these defects was recognized, but their seriousness was not then generally appreciated. It is important that we examine closely the details of American banking in this period in order to understand the causes of a failure without parallel in the history of banking and to appraise intelligently the reforms which have been made in recent years.

Composition of the banking system. Reference has been made to the fact that the United States has always been a nation of small independent banks. The numbers and resources of the various classes of banks which make up the local banking structure, as of June 30, 1921, and June 30, 1929, are shown in the following table, which contains also statistics of membership in the federal reserve

system.¹ Corresponding figures as of March 31, 1938, will be found in the next chapter.

BANKING SYSTEM OF THE UNITED STATES, JUNE 30, 1921 AND 1929

I. All Banks

	<i>Number</i>		<i>Resources (in billions)</i>	
	1921	1929	1921	1929
National banks	8,154	7,536	\$20.5	\$27.4
State banks	18,875	14,437	14.2	16.8
Loan and trust companies . .	1,474	1,608	8.2	16.1
Stock savings banks	978	747	0.5	1.5
Mutual savings banks	623	611	6.0	10.0
Private banks	708	391	.2	0.2
Total	30,812	25,330	\$49.6	\$72.0

II. Federal Reserve Membership

	<i>Number</i>		<i>Resources (in billions)</i>	
	1921	1929	1921	1929
National banks	8,154	7,536	\$20.5	\$27.4
State member banks	1,595	1,171	10.0	18.5
Total member banks	9,749	8,707	\$30.5	\$45.9
Non-member banks	21,063	16,623	\$19.1	\$26.1

Omitting private banks and savings banks, the country's banking business was handled in 1929 by something over 23,500 banks. About a third were national banks, having average resources of about $3\frac{1}{2}$ million dollars and owning slightly over 40 per cent of total bank resources, while the remainder were state banks and trust companies with average resources of around 2 million dollars. This is in marked contrast with the situation in most other countries; in Canada, for example, where in 1928 there were 10 large central banks, with average resources of nearly \$350 millions.

Even these figures do not show as convincingly as do the figures on bank capitalization the extent to which the small bank prevailed.

¹ *Annual Report of the Comptroller of the Currency, 1929.* The figures used in the remainder of this section are also derived from this source.

In 1929 only 28 per cent of the banks of the country had capital of \$100,000 or over; of the remaining 72 per cent, 21.9 per cent had less than \$25,000, 21.5 per cent had exactly \$25,000; with the rest between \$25,000 and \$100,000. Over 80 per cent of the banks were located in towns of 10,000 or less population, but these banks accounted for only 23.4 per cent of the total capitalization.

Branch banking and chain banking. Without doubt the characteristic decentralization of banking in the United States would not have persisted as it has had branch banking been generally permitted. Only in recent years have there been any important steps away from the prevailing condition of local decentralization. In California and certain other states the law has permitted branch banking, and great systems of branches have been developed, state-wide, county-wide, or city-wide. Competition of these state banking institutions finally forced Congress, after a long and bitter legislative struggle, to grant to the national banks certain branch banking powers. The McFadden act of 1927 provided that in any state which permitted branch banking both national and state members of the federal reserve system might, with certain limitations as to national banks, establish and maintain branches within the limits of the city where the parent bank was located.

By 1929 there were 818 banks operating 3,440 branches in the United States, thus making the number of banking offices greater by this number than the total number of banks given in the table above. The great majority of the branch offices were situated in the local communities of the respective parent banks. While these banks with branches formed an important part of the banking system, owning about 40 per cent of the loans and investments of all banks, branch banking had not proceeded extensively enough to alter the essential character of the American local banking system.

Another development of considerable importance was the establishment of elaborate "chains" or "groups" of banks, held together by various devices, the most common being the holding company. This was a device which in fact permitted branch banking, but without the sanction of state or federal law. These bank holding companies were not subject to federal control and

supervision, until made so by the banking act of 1933. In June, 1929, there were 275 such chains or groups of banks in the United States, controlling 1,802 unit banks, which in turn owned 40 per cent of all the branches in the country. The loans and investments of all banks controlled by chains or groups amounted to about 15 per cent of the loans and investments of all banks. Nevertheless important as they were, the chains did not permeate the banking structure sufficiently to be a major factor in the situation.

Weaknesses of decentralized banking. A system of small, independent, local banks contains inherent weaknesses. The multiplication of banks divided minutely the country's banking capital and deposits. Division among so many banks made most of the local banks so small as to defeat effective management and service and cause weaknesses, both of the individual banks and of the system, that were bound in the end to spell disaster. Towns were jealous of one another, and so each had to have its bank. Dissatisfied customers of that bank or ambitious local capitalists or adventurers — spurred on by professional bank organizers, who took a liberal fee or a comfortable official berth or sold banking fixtures — started a second or a third bank in a town that could not decently support one. Banking became one of our most over-expanded industries. Competition between these banks, many in a failing condition, was severe, ruthless, and unintelligent. High interest rates were paid on deposits, and voluminous services were performed without charge or at a loss. Interest rates on loans were frequently exorbitant, by sheer necessity, and loans could be made only locally and to necessitous borrowers at such rates; loans and discounts in the better types of paper in the central money markets were denied by similar necessity, and the investment portfolio had to consist largely of nominally high-yield bonds and mortgages regardless of safety or liquidity. The banks competed in "service to the community," which largely meant buying local securities and making loans on local projects based on sanguine faith rather than intelligent analysis of the borrowers' ability to repay. Thus the bank's portfolio of loans and investments was composed of local items, and the life of the bank was made contingent upon the stability, permanence, and success of the small

locality. Bank buildings were generally disproportionate to the rest of the town in architecture and equipment, and the staff had to be excessive. The salaries, on the other hand, were too small to be attractive to men of adequate training and capacity. Anyone could go into banking with little capital and less of experience and training.

The small town state banks were on the whole indifferent, if not hostile, to the federal reserve system. In 1929 only seven per cent of all the state banks belonged to the system, and these were the larger banks. The national banks in small towns or cities had a difficult time competing with the state banks, for in many states the banking laws were more liberal and the supervision more lax. This made the way of the national bank supervisors difficult, since severity was frequently answered by threats to denationalize and take out a state charter. The difficulty was partly adjusted by lowering the national bank standards (by the act of 1927 and other amendments) as well as by less punctilious examinations.

Such unification as existed was afforded for most banks by a correspondent relationship with a metropolitan bank. The founders of the federal reserve system had indeed cherished the hope that the establishment of the system would mark the end of the decentralization of bankers' balances and the divorce of commercial banking from the stock market, but in actual fact correspondent relationships remained as important as before and bankers' balances became larger than ever. Non-member banks continued to keep their lawful reserves as deposits with metropolitan correspondents, and member banks continued to send surplus funds to banks in large centers. The federal reserve banks were not able to perform all the services formerly done by correspondent banks, through incapacity or legal inhibition or indifference. The failure of the federal reserve banks to pay interest on deposit balances resulted in minimum balances being kept there; the rest was deposited with correspondents, largely in New York City, and loaned in the open discount market or in the call loan market. Correspondent relationships offer at best a poor form of centralization. The correspondent bank is of course legally bound to the

fulfillment of its contractual obligations, but beyond that, particularly in time of stress, nothing more can be expected.

Affiliates. Another defect in the structure of American banking was the close affiliation of commercial banking with other forms of enterprise. Not infrequently close relationships between commercial banks and companies dealing in securities or in real estate were created by the holding company device; in other cases the bank itself owned the controlling stock in an affiliated company. It is the general consensus of opinion that a commercial bank should be divorced from investment business, real estate promotion, and other such interests, especially if it has a pecuniary interest in the profits of the other company. Indeed it is difficult to see how an officer of a bank can bring an unbiased judgment to bear on the wisdom of making a loan to a customer which will enable him to purchase an issue of stock or a piece of real estate offered for sale by a company affiliated with the bank.

Expansion of credit. Thus far we have emphasized the defects in the structure of the American banking system, but there were functional defects as well which diminished the ability of the banks to withstand strains. An exaggerated expansion of bank credit runs like a supporting thread through the web of pre-depression problems in all countries. In the United States it reached the extreme.

The conception of the fathers of the Federal Reserve was that the proper volume and character of credit would be attained by control of rediscounting. Careful definition of eligible paper was counted upon to assure that credit would be extended only to meet the needs of agriculture, industry, and commerce; and they took pains to exclude paper used for speculative or capital purposes. They sought also to reform the credit practices of the commercial world by substituting the two-name acceptance for the single-name promissory note and the open-book account. It was believed that by manipulating the discount rates of the federal reserve banks discounting could be made more or less profitable to the member banks, thus encouraging or discouraging rediscounting and controlling the amount of credit flowing at any time into a particular area or industry.

For several reasons this scheme of control did not work as planned. Changes in the discount rate did not increase or decrease costs sufficiently to influence business activity, and the banks were influenced less by profitability than by their reluctance either to go deeply into debt or to stay continuously in debt to the federal reserve banks. Business men found ways of procuring credit on a short-term basis and using it for long-term capital purposes, and their banks connived with them to this end. The volume of commercial paper, rising with the price level and with business activity, promoted too rapid expansion of credit. The Federal Reserve controlled too small a portion of the banks of the country, and even this control of credit was by indirection, through the reserve banks to the member banks. Furthermore, as has been pointed out, the bars were let down in 1916 and later, making it easier for member banks to borrow against government securities than to rediscount eligible paper.

The difficulties of credit control, inherent in the system, were augmented by the huge inflow of gold from countries off the gold standard or on a gold-exchange standard and by the "easy money" policies of the federal reserve banks. This federal reserve policy was dictated partly by political and industrial conditions in the United States and partly by a program of coöperation between the Federal Reserve Bank of New York and the Bank of England. On the basis of the excess reserves of gold came a great expansion of bank credit. From 1922 to 1929 the loans and investments of commercial banks increased by about fifteen billion dollars, with the total for all member banks of the federal reserve system rising to thirty-six billion dollars.

Most of the increase in credit was neither needed nor used for normal short-term commercial operations but found its way into speculative channels. Brokers' loans and loans to customers to finance the purchase of securities increased at a very rapid rate, and the banks became involved also in financing speculation in real estate and building construction. Inevitably banking assets became less liquid. Loans secured by stock or bonds listed on an exchange are ordinarily not only safe but quite liquid, but they are not rediscountable with the federal reserve banks, and in the

event of a crash on the stock exchange they may be, as the fact proved, extremely difficult to liquidate without loss. Loans on real estate are notoriously difficult to liquidate, and the crisis of 1929 found many banks possessed of mortgages on real estate which they could neither collect nor sell or of real estate which could be sold only at ruinous loss. The expansion of bank credit made possible also an extension of installment financing, and this brought into the banks' portfolios paper which might be sound enough in the long run, but which at this time was not eligible for rediscount with the federal reserve banks nor saleable to other banks in time of crisis.

Not only did bank assets become less liquid than they had been; there was also deterioration in quality. This is reflected both in the inferior security pledged as collateral in many cases and in the increase in the volume of personal loans — unsecured loans made by the directors and officers to themselves or to their personal friends. With both general impairment in the quality and over-extension in the quantity of bank credit, loss of confidence, price reductions, and decline of business activity were sure to result when once the onward march of speculative prosperity was halted.

Foreign loans. Special importance attaches to the character and volume of the foreign loans taken up in the United States in this period of generally expanding credit. Investment bankers were finding a public with ample bank funds and eager for speculative ventures, and they served this public with multiplying high-commission issues of the securities of foreign governmental units and of foreign corporate enterprises. The annual net long-term foreign investments of the United States rose to huge volume by 1928 and then started to decline, although the decline was partly compensated by an increasing flow of short-term credits abroad. There was also a large direct American investment abroad in the acquisition of the capital assets of mining enterprises, chain stores, automobile plants, electrical enterprises, and so forth. At the end of 1932, after considerable reductions from both default and resale, the private foreign investments of the United States amounted, at conservative estimate, to fifteen billion dollars, of which seven

billions were represented by security investments in the ordinary sense and eight billions by direct ownership of capital assets.

Stimulation of exports. The annual lending of immense sums abroad had permitted the merchandise exports of the United States to expand, in spite of rising trade barriers and against the influence of net payments due the United States because of its creditor position. This made possible large exports of heavy mechanical equipment and of agricultural products, serving to maintain these types of enterprise in precarious balance, despite the overexpanded condition of their productive capacities in relation both to the domestic market and to the reviving and expanding productive capacities of other parts of the world. The new annual loans also served to conceal the inherent difficulties — both as to budget problem and transfer problem — of intergovernmental debt and reparation payments in a world of rising trade barriers and badly distributed monetary stocks of gold. The loans to the debtor nations enabled them to postpone resort to increased taxation to provide budgetary surpluses for payments on war indebtedness. Moreover the proceeds of the loans made immediately available foreign exchange which could not otherwise have been obtained except by building up balances of merchandise exports, a thing which the whole trading world was bending every effort to prevent by means of restrictions on imports. Such financing of current requirements out of capital commitments is necessarily a self-limiting process.

Gold exchange standard. The growing foreign investment of the United States also helped the establishment of the gold exchange standard, which became quite prevalent between 1925 and 1929, under which system balances in foreign banks were substituted for actual gold as currency reserves. This practice permitted a much greater expansion of credit over the world than would have been possible if each central bank had held its own actual gold reserves, since both the actual gold in one country and the claims in another country against that gold were counted as reserves. The serious weakness of a generalized gold-exchange standard became painfully evident when the overextended structure of international short-term credit finally collapsed. By 1929 there were a billion

dollars due from foreigners on short-term capital account in the form of deposits, advances, short-term loans, and acceptance credits, although all but 150 millions of this credit was offset by foreign short-term balances held in the United States.

The financial crises of 1929 and 1931. Despite persistent depression in agriculture, industry in the United States had with minor exceptions been fairly prosperous since 1920, but in 1929 the tide turned. Funds in the United States were increasingly in demand for speculation in the domestic security markets. It was becoming evident that the structure of international finance was unsound. With foreign lending reduced, the export industries of the United States were either losing their foreign markets or saving some of them by taking new credit risks and selling on open account; and the unnatural export balance of the United States was declining. By the summer of 1929 most lines of business activity in the United States were showing some recession, and this was fatal to a domestic structure of speculative credit which was based on expectation of everlasting expansion, of accelerated profit taking, and of security values which could move up only. Liquidation made a sudden start on its predestined long journey, foreign lending came to a standstill, and the postponed and now intensified problems of post-war economic readjustments presented themselves for fresh consideration.

In the summer of 1931 a crisis was reached in international financial relations, with successive breakdowns in Austria, Germany, and Great Britain; this was despite considerable international coöperation, in which the United States participated by moratoria on the allied war debts and on commercial debts, as well as by special advances to foreign central banks. Great Britain left the gold standard in September, 1931, and was followed by upwards of thirty countries, leaving a slender "gold bloc" consisting of France, Holland, Belgium, Switzerland, and Poland. There was universal withdrawal of balances of short-term capital from foreign banks, especially from the United States. Among the consequences of this action were renewed credit liquidation, heavy cash withdrawals from the banks, and increased failures in the United States.

Bank failures. The following table shows the numbers of bank failures in the United States from 1921 through 1932. It is the worst record of bank failures in the world's history, despite the fact that America was the wealthiest nation in the world and had gained most and suffered least from the World War. With a sound banking system this crash would have been utterly unnecessary.

BANK SUSPENSIONS IN THE UNITED STATES¹

<i>Year</i>	<i>National banks</i>	<i>State mem- ber banks</i>	<i>Total mem- ber banks</i>	<i>State non- member banks</i>	<i>All banks</i>
1921	52	19	71	434	505
1922	49	13	62	305	367
1923	90	32	122	524	646
1924	122	38	160	615	775
1925	118	28	146	472	618
1926	123	35	158	818	976
1927	91	31	122	547	669
1928	57	16	73	426	499
1929	64	17	81	578	659
1930	161	27	188	1,164	1,352
1931	409	107	516	1,778	2,294
1932	276	55	331	1,125	1,456
Total, 12 years	1,612	418	2,030	8,786	10,816

Neither England nor Canada was completely insulated from the influence of the forces we have been describing, yet not a single bank failed in England on account of the depression, nor in Canada.

Analysis of the bank suspensions since 1921 shows that the rate of failure was higher among non-member than member banks, among banks in the smaller ranges of capitalization, among banks in the towns and cities of small population, among banks in agricultural areas as compared with industrial territory, and among state banks as compared with national banks because of smaller size and looser supervision.

The bank failures since the beginning of the depression of 1929-36

¹ *20th Annual Report*, Federal Reserve Board, 1933, p. 206.

Bank failures since March 4, 1933, lack statistical comparability with those of the 1921-32 period. President Roosevelt's earliest act was to close all the banks in a nationwide "bank holiday"; and when the panic had passed only the stronger ones (ostensibly at least) were licensed to open, and the weak banks were licensed only after they had reordered their capital structures and satisfied the authorities that they were sound. The non-licensed banks were placed in liquidation or receivership, or absorbed, or succeeded by other banks. Banks that have never opened or have been liquidated or put into receivership or have been taken over by stronger banks might well be considered as failures and put into the failure table.

have been more evenly distributed over the country, since the industrial areas were now as badly affected as the agricultural areas or even worse, and the investment portfolios of all banks were disastrously impaired by the radical decline in the values of securities and real estate.

XXVI

AMERICAN BANKING SINCE 1929

Banking legislation since 1929. Great importance attaches to the efforts of Congress and the state legislatures since 1929 to mend the weaknesses of the banking system which were demonstrated by the stock market boom of 1926-29, the panics of 1929, 1931, and 1933, and the depression of 1929-36. This legislation, especially that of the federal government, not only amends the structure and the practice of the Federal Reserve and the member banks but also puts into effect new theoretical conceptions of the nature and functions of central banks.

The federal legislation of this period had six general objectives: (1) to meet the banking emergency wrought by the panic; (2) to promote recovery from the depression; (3) to strengthen the hand of the Board of Governors and the federal reserve banks in obviating or restraining excessive booms in the securities markets like that which broke in 1929; (4) to centralize authority and responsibility and to increase the capacity of the Board of Governors to control money and credit; (5) to improve the banking structure and the practice of member banks, with a view to greater stability and better service to agriculture, industry, and commerce; and (6) to fill certain gaps in the national financial organization which, in the opinion of the administration, private initiative had left unfilled.

Measures taken to meet the emergency we shall pass over briefly, except as to certain actions that promise to have a permanent effect on the banking system. Mention has already been made of some of these actions, such as the closing and reopening of the banks and the sequestration and expropriation of the nation's gold stock.

Pre-"New Deal" remedies: National Credit Corporation and Reconstruction Finance Corporation. The increasingly critical condition of American banks, evidenced by the failure of 1,072 banks in the five months beginning with September, 1931, and the intensi-

fication of hoarding, led the United States government to take various measures to support the credit situation. The first of these was the organization in October of the National Credit Corporation for self-help among the banks. This involved the pooling of resources so that loans might be made on call to distressed members. The National Credit Corporation did not however prove to be of much service, and in February, 1932, it was replaced by the newly created Reconstruction Finance Corporation.

The RFC had an initial capital of 500 million dollars, subscribed by the government, and was authorized to borrow on its own obligations up to three times the amount of its capital. It was empowered to make loans to banks, building and loan associations, insurance companies, mortgage companies, federal agricultural banks, agricultural and livestock credit corporations, and to the railroads and farmers. Later in the year the lending power of the corporation was broadened to permit loans for self-liquidating public works undertaken by states, municipalities, and other public or semi-public agencies. The central purpose of the corporation was to use government credit for thawing frozen private credit. Of similar purpose were the appropriation of 125 million dollars for additions to the capital of federal land banks, operating in the farm mortgage field, and the creation with government capital of a system of home loan banks as central discount and reserve institutions for financial companies operating in the field of urban mortgages. The corporation's capital and borrowings were increased greatly under the Roosevelt régime, for it was utilized to supply the capital funds needed to reopen the banks and to prepare them for deposit insurance, to facilitate the liquidation of failed banks, to enable the Home Owners' Loan Corporation to alleviate home owners distressed by overdue taxes, interest, and principal payments, and to start the NRA, the CCC, and many other agencies of the New Deal. The giant Reconstruction Finance Corporation promises to play a prominent rôle in both long-term and short-term finance for many years to come.

Glass-Steagall act. One other phase of government policy in relation to credit conditions was developed. Early in 1932, as a result of continuing large withdrawals of foreign balances in gold,

with most of the countries of the world already off the gold standard, there was considerable discussion of danger to the gold standard in the United States. The Glass-Steagall act of February, 1932, was proclaimed as a safeguard against such danger. This danger actually did not exist. Withdrawal of all foreign balances in the United States and the exercise of all short-term claims against the United States could not have forced the country off the gold standard, because the tremendous gold stock of the country would not have been dangerously depleted. The gold stock would, it is true, have been reduced to such an extent that the federal reserve banks might have been forced to reverse their "easy money" policies, compel a contraction of credit, and retire federal reserve notes from circulation. In the absence of sufficient holding of eligible commercial paper for use as 60 per cent backing for federal reserve notes, along with 40 per cent gold, the federal reserve banks had been forced to substitute considerable amounts of gold in excess of the minimum requirement of 40 per cent in the security for these notes. The excess gold reserves of the federal reserve banks, calculated on the basis of the 40 per cent minimum gold backing for notes, were not therefore entirely free for withdrawal and export or hoarding; they were "free" only to the extent that they were not needed for this extra backing for the federal reserve notes in place of commercial paper.

The Glass-Steagall act permitted advances by federal reserve banks to member banks on paper usually ineligible and permitted substitution of government obligations for commercial paper for that amount of gold (above the 40 per cent minimum gold requirement) used as the backing of federal reserve notes. The act by the latter provision added 750 million dollars to the "free gold" of the federal reserve system. The federal reserve banks were now in position, without fear of gold withdrawals, to pursue an intensified "easy money" policy, and they did so with a program of heavy purchases of government obligations.

The act was understood at the time to be an emergency measure, and its operation was limited to one year. It has however been extended by successive biennial enactments to 1939. The making of government securities eligible as collateral for the issue of federal

reserve notes has, as noted in a previous chapter,¹ changed the basic character of the notes, and permitting advances by the reserve banks to member banks on any paper acceptable to the reserve banks carries to the limit the distortion of the Federal Reserve founders' conception of a system based upon the rediscount of short-term self-liquidating paper, a distortion started in 1916 when the reserve banks were authorized to make advances to members against government securities.

Banking collapse. The general aim of the depression policy of the government up to the end of 1932 may be said to have been to provide expanded resources of credit so that prospective borrowers might find means available for undertaking a renewal of economic activity. The difficulty with the policy was that prospective borrowers, for the most part, had no apparent incentive to become actual borrowers and so put the increased credit resources to work. The policy was more effective in its negative aspect of preventing a dangerously rapid and possibly excessive liquidation of credit.

The creation of the Reconstruction Finance Corporation and the enactment of the Glass-Steagall law strengthened the public confidence; the number of bank failures declined, the hoarding and exportation of gold were reversed, and in the summer of 1932 the indexes of business turned upward throughout the world.

This progress however proved impossible to maintain, largely because of inherent weaknesses in the banking system. In the late fall and winter came knowledge of the accumulation of a deficit of more than three billions in the federal treasury in the last two years, the renewal of gold exportation, the publication of the exceedingly large amount of loans made by the Reconstruction Finance Corporation to banks, and the November elections, which brought a new party into power whose policies were not fully predictable. The situation grew rapidly worse throughout the country, culminating in a general banking collapse in the first week of March.

Emergency banking legislation: Original act. The first act of the "New Deal," which came into being in March of 1933, was to clear up the banking situation. By March 4 most of the banks of

¹ Chapter XXV.

the country were temporarily closed under state decrees. On March 5 the President, under authority of a war-time statute relating to the control of foreign exchange and the export and hoarding of gold, ordered all banks closed for four days and all banking transactions suspended and placed an embargo on gold withdrawals and gold exports. Congress was called in special session on March 9 and on that day passed the emergency bank act, giving the President emergency powers over the banking system. The bank closing was thereupon indefinitely extended in order to give the Secretary of the Treasury time to classify the banks as to their condition of fitness for reopening.

The emergency bank act gave the President control of all transactions in foreign exchange in any national emergency, gave the Secretary of the Treasury power to call in all gold and to impose penalties for non-compliance, and put the operations of member banks in the federal reserve system under the prescriptions of the President and the Secretary of the Treasury in any emergency. The act was however concerned chiefly with provisions for facilitating the reopening, reorganization, or regulated liquidation of the closed banks. What was done in this regard has been described in the last preceding chapter.

To allay the fears of a money panic caused by the ever increasing hoarding, the federal reserve banks were given power to issue federal reserve bank notes against government obligations (up to the face value) or against notes, drafts, bills of exchange, and bankers' acceptances (up to 90 per cent of the face value) deposited with the federal reserve bank as security. This was distinctly a temporary expedient, for it was provided that, when the President should by proclamation give notice that the emergency had ended, no further federal reserve bank notes should be issued unless backed according to the original plan by government bonds having the circulation privilege, defined by an amendment to the home loan bank act of 1932 as consisting up to 1935 of bonds bearing an interest rate of not more than three and three-eighths per cent and thereafter of two per cent bonds.

Under the supervision of the Secretary of the Treasury and the Comptroller of the Currency banks in need of new capital were

enabled to sell cumulative, preferred stock, without double liability, to the Reconstruction Finance Corporation, whose obligations for this purpose might be increased without limit. The RFC in turn might offer such stock for sale in the open market. Banks prohibited by state law from issuing such stock were permitted to sell debentures or capital notes instead. Under this authority the corporation had, by the end of 1935, purchased \$832,852,934 of preferred stock from 4,134 banks and \$435,177,780 of capital notes and debentures from 2,847 banks. This investment, totalling somewhat over a billion and a quarter dollars, had been reduced to \$559,625,000 by December 31, 1937.

These purchases improved the capital structure of many banks and enabled them to reopen, to qualify for membership in the Federal Deposit Insurance Corporation, and to lend in larger amount and with less hazard. On the other hand, this development was viewed with considerable apprehension by many citizens, who feared that this entry of the government into banking might interfere with private finance and presage the ultimate socialization of the banks.

Perpetuation of emergency legislation. The Roosevelt administration has been reluctant to concede that the emergency has passed and has been willing to retain in force to the present time many of the measures adopted for meeting the emergencies of 1932-33. Some of the measures stand a good chance of becoming permanent features of the banking system, especially as certain misgivings have weakened or vanished through actual experience under the emergency. For instance the insistence on keeping the federal reserve system exclusively a commercial-credit system has well-nigh disappeared.

Another example of this change of attitude pertains to a provision for meeting future panics. The Glass-Steagall act of 1932 was an emergency measure, devised to remedy a defect in the capacity of the Federal Reserve to meet the necessities of intense crises. The banking act of 1935, following the precedent of the Glass-Steagall act, provides that a federal reserve bank, under rules and regulations prescribed by the Board of Governors of the Federal Reserve System, may make advances to a member bank on its

time and demand notes which have maturities of not more than four months and which are secured to the satisfaction of the federal bank. Each such note must bear interest at a rate not less than $\frac{1}{2}$ per cent higher than the highest discount rate in effect at the federal reserve bank on the date of the note.

This authority is a continuing privilege to the member bank to use any of its assets as security for a loan from the federal reserve bank at any time, provided that it can satisfy the reserve bank of the desirability of the loan and the soundness of the collateral and is willing to pay the penalty rate on such loan. Such a privilege is consonant with the practice of the Bank of England, which makes lombard loans (*i.e.*, loans based on securities as collateral) at rates above its "bank rate" on prime paper, although it is inconsistent with the conception of the founders of the Federal Reserve.

A third illustration is the Reconstruction Finance Corporation, which, with its thirty branches throughout the country, is in reality a huge government bank, organized for an emergency, but easily adaptable if desired to the rôle of a continuing competitor of the private banking system. Created in 1932 as a one-year emergency organization, the corporation's life has been extended several times, and it is difficult to see how it can be dispensed with for a decade or more. True its predecessor, the War Finance Corporation, was extinguished within a decade, but it was conceived as a war measure, and the war was indisputably past. The Reconstruction Finance Corporation, on the other hand, was created as a peacetime institution, and it has come to undertake, not merely emergency duties, but also functions that may continue into the indefinite future. It has also become intimately, if not inextricably, articulated with the Treasury and the financial system of the country. The renewal and extension of services during the depression of 1937-38 indicates that it may be kept to handle future emergencies and to promote the policies of the government. On December 31, 1937, the total outstanding loans and investments of the corporation were \$2,017 millions, of which \$290 millions were loans to or investments in the capital stock of other governmental financial institutions, such as the Commodity Credit Corporation, the RFC

Mortgage Company, the Export-Import Banks, and the Rural Electrification Administration.

A plain instance of an emergency measure becoming a permanent fixture as a device for handling future emergencies is the so-called "direct industrial loans." In the fall of 1933 the Reconstruction Finance Corporation announced its readiness to lend up to one billion dollars at 3 per cent to banks which would agree to relend to business at a rate no higher than 5 per cent on six months' maturity for the purposes of buying materials, covering labor costs in manufacturing, or helping merchants especially affected by the program of the National Recovery Administration. The Reconstruction Finance Corporation stood ready to give the banks advance rulings on the acceptability as collateral of slow assets, which commercial banks would not ordinarily accept.

This emergency measure, although of slight importance in actual application at the time, technically put the Reconstruction Finance Corporation in direct competition with the federal reserve banks in rediscounting. Further competition of the same type, together with the entry of these two agencies into competition with the commercial banks, was embodied in the legislation of June, 1934, which authorized the federal reserve banks, to the extent of 280 million dollars, and the RFC, to the extent of 300 million dollars, to make five-year loans for working capital directly to industrial corporations "unable to obtain requisite financial assistance on a reasonable basis from the usual sources." On December 31, 1937, the RFC had outstanding 75 million dollars of such loans and the federal reserve banks 18 millions. The head of the RFC has repeatedly declared that his institution would not retire from competition with the banks until they became active and took over this type of lending, repeatedly threatening that if they did not loosen up the government (presumably through his corporation) would be forced to take over and operate the banks. As a matter of fact the banks have generally been most eager to extend their lending activity, drastically reduced as it was by industrial depression, and have been deterred only by the absence of borrowers presenting sound propositions. The government's attitude appears to bring pressure upon the banks to make unsound loans.

Structural reforms of 1933 and 1935: Deposit insurance. In addition to emergency features, the banking legislation of 1933 and 1935 contained many important provisions intended to bring about permanent reforms in the banking system. We may examine the more important of these under the two heads of structural reforms and functional remedies.

Prominent in the former group was the establishment of a nationwide system of insuring bank deposits. Though growing directly out of the banking emergency of 1933, this device was heralded as a preventive of similar panics in the future. The system established is not a government guaranty of deposits but the self-insurance of their deposits by the banks themselves. Every national or state member bank is required to be an insured bank subject to the provisions of the law. The plan is open to non-member banks, but no state bank which during the calendar year 1941 or any succeeding year has average deposits of \$1,000,000 or more may be an insured bank or continue to have any of its deposits insured after July 1 of the year following any such calendar year, unless the bank is a member of the Federal Reserve.

The Federal Deposit Insurance Corporation was authorized at its discretion to operate a separate fund for the benefit of the mutual savings banks and their depositors. Conforming to the authority, such a fund is now operated in a wholly separate manner for the mutuals alone, and the capital assets of the corporation, as well as all its expenses, are allocated between the funds on an equitable basis. Such separate treatment of mutual savings banks was warranted by the fact that their deposits are all time deposits and that the mutuals operate, as experience shows, with a very low loss ratio. The scheme has not been popular with the mutuals; the corporation in 1936, for instance, was insuring the deposits of only 56 such banks.

The maximum amount insured for any depositor in one bank is \$5,000. In applying this limit, all balances standing to the credit of a depositor in a single bank are combined. The assessment rate is one-twelfth of one per cent per annum for commercial banks, applied to the average net deposits for the preceding six months of the calendar year. Any national bank failing at any time to comply

with the assessment requirements automatically loses all its rights, privileges, and franchises under the national bank act or under the federal reserve act, and no insured bank may pay dividends on its stock while it is in default in the payment of any assessment. The corporation, after giving due notice that an insured bank or its directors or trustees have continued unsound or unsafe practices in conducting the business of the bank or have knowingly or negligently permitted any of its officers or agents to violate any of the provisions of the law or regulations to which the insured bank is subject, may, after a hearing, suspend the insured bank from the insurance protection and publish this fact. Any insured bank, except a national bank or state member bank, may voluntarily withdraw from the insurance plan upon not less than 30 days' notice to the corporation, with notice also to the Reconstruction Finance Corporation if the latter owns or holds as pledge any of the bank's preferred stock, capital notes, or debentures. The insurance of the deposits, as of date of suspension or withdrawal, continues two years thereafter, but no new deposits are insured. Whenever a bank ceases to be a member of the federal reserve system, its insurance stops forthwith.

Whenever an insured bank is closed because of inability to meet the demands of its depositors, the Comptroller of the Currency, if it be a national bank, must appoint the FDIC as receiver for the closed bank. The receiver realizes on the assets of the closed bank, having due regard to the condition of the credit of the community. The corporation is required, upon failure of the bank, to pay the depositors as soon as possible. The corporation may, if it deems it advisable, organize a new bank and transfer to it the assets and liabilities of the old bank, and the capital stock of this new bank may in time be sold to the public, or in default of such sale, the assets and liabilities may be transferred to another insured bank, as the FDIC may see fit.

The corporation must appoint examiners, who shall have power on behalf of the corporation to examine any insured state non-member bank, any state non-member bank making application to become an insured bank, and any closed insured bank, whenever in the judgment of the corporation an examination is necessary.

Each insured state non-member bank must make reports of condition in such form and at such times as the corporation may require.

The total amount of deposits insured by the corporation on May 13, 1936, was estimated to be about 20,500 million dollars, of which 19,600 millions were in the 14,092 insured commercial banks. Total deposits in the insured commercial banks amounted to more than 45 billion dollars, held in over 57 million accounts; total deposits in the 56 mutual savings banks amounted to 980 millions, held in nearly 1,400,000 accounts. Approximately 43 per cent of the deposits in the insured commercial banks and 89 per cent of the deposits in the insured mutual savings banks were estimated to be insured. The total number of banking offices (banks, branches, and additional offices) on December 31, 1936, was 18,516. Of these 17,234 were insured. Of the 50,800 million dollars of total deposits of all commercial banks, 49,200 millions, or 97 per cent, were in insured banks.

Judgment of deposit insurance. The deposit insurance plan has met with severe criticism from bankers and economists, and indeed the experience in the United States with various schemes for the guaranty of bank deposits would seem to warrant misgivings as to its right to a permanent place in the banking structure, however essential it may have been considered in 1933 as a means of overcoming the quite natural timidity of depositors and counteracting the trend towards hoarding. New York had a guaranty plan in 1829 that was bankrupt in 1837. In more recent years there were such plans in operation in Kansas, Mississippi, Nebraska, North Dakota, Oklahoma, South Dakota, Texas, and Washington. In every case there were strong forces tending to lower the standard of quality in banking practice. In every case the plan quickly broke down and was abandoned. South Dakota recently wound up its bank guaranty fund with assets of seven hundred thousand dollars and liabilities of thirty-five millions. The continued operation of the schemes would have put all solvent banks in bankruptcy.

Whether similar results will follow the adoption of the insurance plan by the federal government is of course impossible to predict. In a national scheme there is a much greater diversification of risk

than in a state scheme, where all the banks are frequently dependent on a single type of loan and in danger of all getting into difficulty together if any economic adversity strikes the leading industry of the region. Again it may be expected that the supervision of the banks under the Federal Deposit Insurance Corporation will be stricter than that which existed in most of the states. Rigid standards set up by federal inspectors are expected by some to compensate for the laxness in management which others believe the plan encourages and condones. As accounts above \$5,000 are not insured, the desire for large deposits may still be a potent force in maintaining high standard of management. There are still others who find encouragement in the thought that the plan will drive state banks into the federal reserve system, believing that depositors will shun those banks not covered by deposit insurance and that thereby the banking system as a whole will be unified and strengthened. The banking act of 1935 renders the plan rather more attractive to the banks and consequently makes less likely the withdrawal of a large number of banks of unquestioned stability seeking to avoid the liability of the insurance plan. It is true nevertheless that a number of the strongest banks, especially in New England, have quit the system since August 23, 1935. Deposit insurance at best is a doubtful substitute for a thorough reorganization of the banking system, national and state, such as would make insurance or guaranty plans unnecessary.

The jurisdiction of the Federal Deposit Insurance Corporation offers of course a new avenue of approach to the exercise of a centralized influence over most of the 9,000 state banks which are not members of the federal reserve system. So long as they feel the need of deposit insurance, the failure of these banks to join the Federal Reserve need no longer hinder their actual subjection to many of the regulations of that system. The possible centralizing influence is further strengthened by the opening of membership in the federal reserve system to Morris plan banks and others of the same type and to mutual savings banks that qualify.

Capital requirements. As a step toward the elimination of difficulties arising out of the keen competition in a banking system composed of very many small and frequently undercapitalized

independent banks, recent legislation has altered capital requirements. Minimum capital for national banks has been increased in all categories, becoming \$50,000 for a bank in any place having a population of less than 6,000. State banks are excluded from membership in the Federal Reserve unless the national bank capital requirements are met. The banking act of 1935 further requires that newly organized national banks must have surpluses equal to 20 per cent of their capital stock and that every national bank must build up its surplus out of earnings to a level of 100 per cent of capital stock instead of the 20 per cent level formerly required. In one respect this accumulation of larger surplus is offset by the loss of the security afforded to depositors by the double liability of shareholders, which has now been abandoned. Against this device it was urged (1) that it discouraged local investors from subscribing to bank stock during the depression and so delayed the rehabilitation of the system, (2) that the protection it afforded to depositors had proved rather illusory, since collections from shareholders were small and expensive, and (3) that deposit insurance was more adequate.

Branch banking. Branch banking received modest encouragement in the banking act of 1933 by the grant to national banks of authority to establish and operate branches either in the city or throughout the state in which the bank is located, provided that similar authority to maintain branches is specifically granted to state banks by the laws of the state in question, and subject to severe restrictions as to the population of the place and the size of the bank, intended to limit the establishment of branches in small places or by small banks. A number of states have liberalized their banking laws in consequence of this action, and branch banking promises to spread when the depression lifts. On December 31, 1937, the number of banks maintaining branches or additional offices was 903, of which 194 were national banks, 159 state member banks, and 550 non-member state banks. The number of branches and additional offices was 3,407, those of the national banks being 1,485, of member state banks 994, and of the non-member state banks 928. 1,648 of the branches and additional offices were in the head-office city, and of those outside the head-

office city 705 were in the head-office county and 425 in contiguous counties.

Holding company affiliates. The banking act of 1933 gave some attention to the related problem of chain banking; that is, the problem arising from the control and direction of a group of banks by a holding company. The act requires such a holding company, called a holding company affiliate, to secure a permit from the Board of Governors before it may vote the stock which it owns in a member bank. To obtain the voting permit the holding company must submit to examination like any national bank, must build up assets other than bank stock, and must dissolve any connection with a security selling organization. Member banks are restricted in the making of loans to a holding company affiliate, in the purchase of its obligations, and in the acceptance of such obligations as collateral for loans to others. Of any group of member banks affiliated with a holding company only one may participate in the nominating and election of directors of a federal reserve bank. Some measure of control was thus extended over a form of banking organization which had been open to considerable abuse.

Divorce of commercial and investment banking. There was mention in the last chapter of the fact that the keenly competitive search for banking profits had developed unwise policies by the granting of credit by commercial and investment banking. Commercial banks devoted more and more of their resources to strictly speculative loans, to real estate loans, to investment holdings, and even to dealings in securities. The shift of funds away from strictly commercial banking functions was intensified under the conditions of business depression.

The banking act of 1933 sought to correct these defects. Limits were placed on loans by any member bank to its own officers, loans to security affiliates were forbidden, and the amount of the obligations of such affiliates which might be accepted as collateral on loans was restricted. Another group of provisions was intended to effect a divorce between commercial and investment banking, their marriage having resulted in some undesirable offspring during the speculative boom and the following depression. In this connection, member banks were forbidden to engage in the investment banking

business, a limit was placed on the amount of any one issue of securities that a member bank might hold on its own investment account, and it was stipulated that no member bank might be affiliated in any way with an organization engaged principally in the issue or distribution of securities, and that no officer or director of a member bank might be an officer, director, or manager of any organization engaged primarily in the business of buying, selling, or negotiating securities, unless with permission of the Board of Governors. No enterprise engaged in investment banking was permitted to engage in the business of receiving deposits. The banking act of 1935 slightly relaxed some of these prohibitions and gave the banks a limited privilege of underwriting and dealing in securities.

Functional remedies of 1933 and 1935: Interest on deposits.

We may give our attention now to certain functional remedies contained in the recent banking legislation. Thus the act of 1935 prohibits member banks from paying interest on demand deposits and directs the Board of Governors of the Federal Reserve System to limit by regulation the rate of interest paid by member banks on time and savings deposits. In addition it should be noted that the Federal Deposit Insurance Corporation must by regulation prohibit the payment of interest on demand deposits in insured non-member banks and similarly must limit from time to time the rates of interest paid on time and savings deposits. The corporation and the Board of Governors have coöperated in fixing the same rates in any territory, and they have also secured the coöperation of the state banking departments in setting rates for non-insured banks.

This change in the operation of the banking business is of considerable importance. The interest allowed on deposits was frequently a chief competitive weapon among banks in their attempts to secure an adequate volume of business to provide a fair return on capital. The rendering of free service for customers in many different connections, sometimes not properly banking at all, was another aspect of the same situation, and such services had been sharply limited under the banking code of the National Recovery Administration. The payment of interest on demand deposits cut

heavily into earnings. A relatively small reduction in the interest paid over the last decade could have provided adequate reserves (had such funds been retained for this purpose) to cover all the abnormal losses of the depression. Certainly the reservation of larger earnings for maintenance of capital, during the years of prosperity would have meant a considerable difference for many of the banks which had been closed before 1932 and would have given a healthier banking system. There is no justification for interest on the demand deposits of a commercial bank; such payments merely encourage the keeping of unnecessarily large balances in a form which is most dangerous for the bank.

The prohibition of interest payments on demand deposits was more or less welcomed by the banks at this time; first because it offset in part the cost of insurance premiums under the FDIC; second because the banks had had a very difficult time to earn their costs and dividends during the depression while safe loan opportunities had been so few and yields on investments so low; and third because they had been surcharged with excess funds held idle for liquidity's sake.

Speculative use of credit. Still another group of provisions in the act of 1933 attempts to set limits to the speculative use of credit. The requirement that federal reserve banks shall discount for members without discrimination is changed to say that they may discount for members; the reserve banks are hereafter to watch the use of credit by members, and the Board of Governors may put an embargo on credit to any member found making an undue use of credit for speculative loans. Limitations on the making of loans on security collateral include a prohibition of the practice of placing such loans with brokers or dealers in the capacity of agent for a non-banking organization. Leading bankers have remained skeptical of the appropriateness of these provisions to accomplish adequately their intended purpose. The stock market boom of 1936-37, with the precipitate crash of 1937-38, seems to support this view.

Centralized control of the banking system. As noted in the immediately preceding chapter, the Board of Governors had not up to 1933 attained to that dominance in the banking system which was by many regarded as desirable. Now comes the "New Deal," whose

philosophy and legislation, actuated in part by the emergency, stands for the centralization and aggrandizement of authority, which in the field of finance means the authority of the President, his Secretary of the Treasury, and the Board of Governors.

The changes making for greater integration within the federal reserve system itself were chiefly those which strengthened control by the Federal Reserve Board, or Board of Governors of the Federal Reserve System as it has been called since 1935. This control was fortified by the power of the board to alter reserve requirements, to regulate the payment of interest on time and savings deposits, to exercise supervision over credit for speculative purposes, to establish regulations governing the standards for rediscounts and loans by federal reserve banks to members, to set the rediscount rates of the banks, and to control through the Open-Market Committee all open-market operations and impose its decisions upon the federal reserve banks. The control was further strengthened by the power granted in 1933 to order the removal of any director or officer of a member bank if after warning he continued to violate any banking law or continued "unsafe or unsound practices."

Experience has shown that the establishment of political influence over the banking system and its credit policies is fraught with grave dangers. With greater centralization of authority over the banking system, signs of increased political influence in the exercise of such authority become doubly significant. Indications are not lacking that recent banking legislation, government activity in the specialized banking fields of agricultural and home mortgage credit, and the banking operations of the Reconstruction Finance Corporation are heading in this direction. In other words, the financial machinery of the country has become to an unprecedented degree the vehicle by which the Administration seeks to accomplish its political, social, and economic objectives. And in these connections the Federal Reserve, clothed with new instrumentalities of control and direction, is being employed to accomplish things new to central bank functions as tradition had molded them.

Control of credit: Concept of credit control. The founders of the Federal Reserve sought, by strict definition of paper eligible for rediscount, a *qualitative control* of credit. The record of the failure

to achieve this has been presented. That faith in the possibility of securing qualitative control of credit has not been wholly abandoned is attested by the presence in the acts of 1933 and 1935 of numerous provisions (which we have noted) granting power specifically intended to ensure a higher quality of bank credit. Nevertheless it is true that greater emphasis has been laid in recent years on *quantitative* than on qualitative control. The "New Deal" has seemed less anxious about the direction which credit is to take and more zealous to increase the quantity outstanding, to flush the bank reserves and ease the money market, to enlarge purchasing power and thereby to promote recovery. The philosophy of the Federal Reserve has thus been shifted, in emphasis at least, to quantitative control, exercised chiefly through open-market operations and reserve manipulation.

Open-market operations. Experience has taught that the most effective instrumentality of control, permitting positive action by the authorities rather than passive waiting for requests from member banks for advances or rediscounts, is the open-market purchase and sale of government securities and acceptances. The new importance attached to quantitative control required that the open-market machinery be brought more fully and directly under the centralized control of the Board of Governors. This was accomplished by the banking acts of 1933 and 1935. A Federal Open-Market Committee has been set up, consisting of the members of the Board of Governors of the Federal Reserve System and five representatives of the federal reserve banks. These representatives are elected annually, one each by the boards of directors of specified groups of federal reserve banks designed to give regional representation. The committee adopts and transmits to the federal reserve banks regulations relating to the open-market transactions of such banks. The open-market operations of every federal reserve bank must be in accordance with the direction and regulations of this committee. The time, character, and volume of all transactions in paper eligible for open-market operations must be governed with a view to accommodating commerce and business and with regard to their bearing upon the general credit situation of the country. The securities eligible for open-market operations are extended to

include all those fully guaranteed by the United States government.

Regulation of reserve ratios of member banks. Second in importance among the instrumentalities of quantitative control of credit is regulation of the member banks' reserve ratios; *i.e.*, the percentages of deposits which they must hold as reserve with the federal reserve banks. Under the banking act of 1935 the Board of Governors of the Federal Reserve System may, "in order to prevent injurious credit expansion or contractions," change by regulation the percentage requirements, provided that it cannot set them below 13 per cent, 10 per cent, or 7 per cent of demand deposits, depending upon the location of the bank, and 3 per cent of time deposits and that they cannot be increased to more than twice those amounts. If a bank required to carry a 10 per cent reserve is directed to jump its reserve to 20 per cent, its loaning capacity will be reduced, and a lowering of the reserve percentage later would enhance this capacity. Of course, instead of curtailing loans when the reserve requirement is increased, the bank may sell securities from its portfolio and transfer the proceeds to its reserve account.

This device is even more direct and powerful than open-market operations in creating or absorbing excess reserves; moreover by reducing reserves and forcing members to become indebted to the federal reserve banks, the board can make changes in the rediscount rate more effective in credit regulation. Such control is a new device in central bank technique; how useful it may prove to be is problematical. If the Board of Governors has actual insight into the situation, proper judgment of the timing, and courage and will power, employment of this device may prove effective to stop a credit expansion. It is likely to be much less effective in bringing about a credit expansion at will. Moreover it may well prove to be too dangerous a tool to employ, since raising the reserve requirements might precipitate a contraction of credit that would be panicky. The experience to date has been none too happy, and the Administration has confessed that the increase in required reserves in the spring of 1937 was too hard on the banks, forcing them to dump government securities, and thus tending to break the market and initiate the depression of 1937-38.

Loans on real estate. Loss of interest in qualitative control is further indicated by a provision of the banking act of 1935 which greatly extends the authority of national banks to loan on real estate security. This represents a departure from the notion of keeping these banks strictly commercial. A national bank is now authorized to make loans secured by first lien upon improved real estate in amount not exceeding 50 per cent of the value of the real estate; the maximum period is five years, except for amortized loans, which may run for ten years if the installments are such as will amortize at least 40 per cent of the principal of the loan within the ten-year period. These restrictions do not apply to insured real estate loans made under the provisions of Title II of the national housing act. The aggregate amount of real estate loans may not exceed the capital and surplus of the bank or 60 per cent of the amount of its time and savings deposits, whichever is the greater.

The relaxation of restrictions on the making of real estate loans was an act of doubtful wisdom. Such loans are notoriously lacking in liquidity and are in consequence a type of investment which the authorities generally agree the commercial banks should avoid. The wholesale bank suspensions and other banking difficulties in the United States in recent years were aggravated by the unwarranted extent to which many banks had diverted funds to loans on real estate. Between the middle of 1918 and the middle of 1929 member banks increased their real estate loans by nearly 600 per cent, while their total loans and investments were increasing by a little less than 100 per cent. At the latter date real estate loans constituted 9 per cent of the total loans and investments of these banks. To encourage still further the making of such loans is definitely to weaken the commercial banking structure.

The present banking situation. The foregoing survey of recent banking legislation shows that some of the weaknesses of the American banking system described in the immediately preceding chapter have been remedied but that at the same time certain new elements of weakness have been introduced. The table on the following page shows the composition of the banking system as a whole as of March 31, 1938.

BANKING SYSTEM OF THE UNITED STATES, MARCH 31, 1938

	<i>Number</i>	<i>Loans and investments</i>	<i>Deposits*</i>
I. All banks			
National banks	5,250	\$20,403,000,000	\$22,264,000,000
State banks :†			
Commercial banks and trust companies	9,459	17,268,000,000	18,753,000,000
Mutual savings banks	563	10,196,000,000	10,259,000,000
Private banks‡	76	452,000,000	421,000,000
Total, state banks	<u>10,098</u>	<u>27,916,000,000</u>	<u>29,433,000,000</u>
Total, all banks	15,348	48,319,000,000	51,697,000,000
II. Federal Reserve membership			
National banks	5,250	20,403,000,000	22,264,000,000
State member banks	<u>1,085</u>	<u>11,118,000,000</u>	<u>11,854,000,000</u>
Total member banks	<u>6,335</u>	<u>31,521,000,000</u>	<u>34,118,000,000</u>
Non-member banks	9,013	16,798,000,000	17,579,000,000

* Excluding interbank deposits

† Comprises all licensed state commercial banks, trust companies, mutual and stock savings banks, and such private and industrial banks as are included in abstracts issued by state banking departments.

‡ Private banks that agreed to examination by the Comptroller of the Currency or a federal reserve bank, under provisions of the act of 1933.

While the total number of banks has been reduced from 30,104 in 1929 to 15,569 in 1938, there still remains that condition of many independent local banks of small average resources which has always represented a fundamental weakness in the American banking system. There remains the imperative need of reduction in the number and increase in the average size of the commercial banks. These results could be brought about either through consolidation of competing banks or through a thoroughgoing extension of branch banking. Some form of branch banking on a wide regional or national basis would appear to be both necessary and inevitable, and certain progress in this direction has already been noted. Without doubt the competitive chartering of banks by both federal and state governments should give place to a uniform national system. At the head of such a system should stand, as now, the Federal Reserve; possibly it could be more efficient if organized as one central bank with branches. Certainly the Federal Reserve should be removed so far as practicable from political influence to the end that it may serve the commercial banks and the national treasury, in a manner illustrated by the Bank of England.

Statement of the federal reserve banks. The nature of the business of the federal reserve system and its present condition, as affected by the many important changes in banking legislation described in this chapter, may be illustrated by the combined statements of the federal banks as presented on the following page.

Other forms of government banking: Agricultural credit agencies. We may conclude this survey by examining certain collateral banking functions which have been undertaken by the federal government. In the special banking field of agricultural credit the federal government has also directly assumed new or amplified banking functions. The various agricultural agencies were consolidated in a new Farm Credit Administration under a single Governor, for the purpose of improving the efficiency of the credit service. The Administration, in addition to the Governor, consists of the Land Bank Commissioner and the Intermediate Credit Commissioner, each of whom heads a centralized system of credit institutions. The institutions thus consolidated consist of (1) the twelve federal land banks — long-term mortgage banks, owned by some 3,000 coöperative national farm loan associations scattered over the country; (2) the joint stock land banks — long-term mortgage banks, similar to the federal land banks, but privately owned and operated and to be liquidated, no new loans being permitted; (3) the twelve federal intermediate credit banks — operating in the same districts and housed in the same buildings and governed by the same directorate as the federal land banks, but lending for periods from one to three years on agricultural products and livestock; (4) twelve production credit corporations, supplied with government funds and authorized to invest in the stock of, make loans to, and in large measure control the activities of local production credit associations, which lend to farmers for general agricultural purposes; (5) the Central Bank for Coöperatives and twelve regional banks for coöperatives, with funds largely supplied by the government, authorized to make loans to coöperative associations which promote the orderly marketing of farm crops by holding, storing, processing, and marketing farm products and by making loans to their members on the security of farm products deposited with them for eventual marketing.

COMBINED STATEMENT OF THE FEDERAL RESERVE BANKS, SEPTEMBER 21, 1938

Assets

Gold certificates on hand and due from U. S. Treasury	\$10,719,741,000
Redemption fund — federal reserve notes	8,690,000
Other cash	383,339,000
Total reserves	<u>\$11,111,770,000</u>
Bills discounted:	
Secured by U. S. Government obligations directly or fully guaranteed	\$5,506,000
Other bills discounted	2,706,000
Total bills discounted	<u>\$8,212,000</u>
Bills bought in open market	\$540,000
Industrial advances	15,683,000
U. S. Government securities	788,655,000
Treasury notes	1,165,205,000
Treasury bills	610,155,000
Total U. S. Government securities	<u>\$2,564,015,000</u>
Total bills and securities	2,588,450,000
Due from foreign banks	181,000
Federal reserve notes of other banks	27,031,000
Uncollected items	610,821,000
Bank premises	44,405,000
Other assets	46,715,000
Total assets	<u>\$14,429,373,000</u>

Liabilities

Federal reserve notes in actual circulation	\$4,201,169,000
Deposits:	
Member bank — reserve account	8,013,536,000
U. S. Treasurer — general account	917,078,000
Foreign bank	195,499,000
Other deposits	146,010,000
Total deposits	<u>\$9,272,123,000</u>
Deferred availability items	603,701,000
Capital paid in	133,998,000
Surplus	175,422,000
All other liabilities	42,960,000
Total liabilities	<u>\$14,429,373,000</u>
Contingent liability on bills purchased for foreign correspondents	234,000
Commitments to make industrial advances	<u>13,553,000</u>

Federal Reserve Note Statement

Federal reserve notes:	
Issued to federal reserve banks by federal reserve agents	\$4,514,135,000
Held by federal reserve bank	312,966,000
In actual circulation	<u>4,201,169,000</u>
Collateral held by agent as security for notes issued to bank:	
Gold certificates on hand and due from U. S. Treasury	\$4,604,000,000
Eligible paper	7,373,000
Total collateral	<u>\$4,611,373,000</u>

The federal government is thus prepared to perform functions which in times past were considered to be exclusively within the province of the banks, but which were not satisfactorily performed by them during the long period of bank closings and liquidations in agricultural areas, if ever. It seems that much of the financing of agricultural operations — of long, short, and intermediate maturities — is permanently taken from the state and national banks and the former loan institutions.

The Congress went still further and created the farm mortgage corporations, to issue and use two billions of dollars of refinancing bonds guaranteed by the government as to both principal and interest. The funds of the corporations are used (1) to refund bonds which had already been issued by the federal land banks for the purpose of refinancing farm mortgages, thus exchanging new bonds fully guaranteed for the former bonds, which were guaranteed only as to interest, and (2) to refinance mortgages of additional mortgagors who might be induced to exchange mortgages for the newer type bonds.

Loans to home owners. The home loan mortgage business had never been adequately provided with funds or effective machinery for granting, servicing, or rediscounting loans. One significant effect of the failures in 1929 and after of the state and national commercial banks, as well as of the mortgage bond and mortgage guaranty companies, following as they did the excessive real estate developments between 1922 and 1929, was the complete prostration of the construction industry and the collapse of real estate values. It was generally felt that recovery from the depression was defeated, if not made impossible, unless this situation were relieved. Into this vent stepped the government, with a series of credit institutions :

(1) The twelve federal home loan banks, established in 1932, as reserve or discount institutions. These are owned by mortgage companies, the federal savings and loan associations, the state building and loan associations and other lending institutions, and the United States government. The basic function of these banks is to help the members finance their lending operations, especially in emergencies.

(2) The Home Owners Loan Corporation. In order to save owners of small homes from losing their equities in their homes, where that danger might exist, and in part to liquidate mortgages tied up in closed banks, a program of refinancing with government funds, similar to the program for farm mortgages, was applied to the field of urban home mortgages by the home owners loan act of June, 1933. A Home Owners Loan Corporation was established with 200 million dollars of government capital and authority to issue (with later increases by amendment) a total of $4\frac{3}{4}$ billion dollars of bonds with interest and principal guaranteed by the government.

The original act authorized the corporation, in cases where the creditor was pressing the debtor, to issue bonds to mortgage holders in exchange for mortgages recorded prior to June 13, 1933, on bona fide homes of a value of \$20,000 or less. Supplemental cash loans were also made when needed for such items as accrued taxes and repairs, and to a limited extent the corporation might also refinance in cash or make cash loans on unencumbered homes.

The act also provided for the amortization of the refunding mortgage, with interest at one per cent more than on the refinancing bonds, over a period of fifteen years, with a three-year extension on payments of principal if desired. The operations of the corporation were viewed as strictly emergency financing, and the corporation discontinued the taking of new loans in 1936. It is now contracting its activities and organization but promises to continue in existence for a considerable time and incidentally to lose no small sum before it is liquidated.

(3) The federal savings and loan associations, organized under and holding membership in the federal home loan bank system, and doing business similar to the building and loan associations. The associations are under federal charter and have capital subscribed in part by the government, but they are under local management. They are intended to fill a permanent place in areas not already adequately served by private mortgage companies, and there is a strong current of conversion of state building and loan associations to federal charters.

(4) The Savings and Loan Insurance Corporation, created to

insure the shares in the federal savings and loan associations and other members of the federal home loan banks, after the manner in which the deposits in the commercial banks are insured by the Federal Deposit Insurance Corporation.

(5) The Mutual Mortgage Insurance Fund, which under the Federal Housing Administration establishes a system of mutual mortgage insurance whereby certain mortgage investments, in the hands of approved mortgagees, may be granted a substantial insurance against loss through default of payments of interest or principal. These are long-term loans, commonly called "Title II loans." The National Housing Act of June 27, 1934, also provided a scheme of insurance of credit employed in the renovation and modernization of homes. These loans are spoken of as "Title I loans"; they are of short term and designed to stimulate building and help end the depression. The legal provision for the insurance of such loans expired in 1937 for a time but was revived in January, 1938, because of the return of depression.

(6) The national housing act provided for the establishment of national mortgage associations, under federal charter, with power to buy and sell first mortgages and to borrow money for such purposes by the issuance of securities. They were designed to absorb the insured Title II mortgages in case the public should not readily buy them. One such institution, the Federal National Mortgage Association of Washington, D. C., was created by the RFC in 1938, and others may follow.

Criticism of government banking. To a certain extent specialized banking activities of the government may be justified as necessary salvage operations in an emergency characterized by general timidity and frequently paralyzing losses among private credit agencies. A great deal of hardship for the families of farmers and of urban home owners has been averted or ameliorated. At the same time it must be recognized that putting government capital into competition with private capital on an extended scale does not tend to restore the active confidence of the latter in the prospects for profitable operations. Officials of the "New Deal" have frequently said that the government wishes to get out of banking activities as soon as may be justified by indications that bankers

are prepared to "resume their responsibility." But the history of the assumption of functions by government suggests the possibility that much of the operation of specialized banking services by the government of the United States may be destined to continue indefinitely, a prospect which is scarcely calculated to inspire confidence on the part of private lending agencies.

XXVII

FOREIGN BANKING SYSTEMS

Banking principles and banking systems. Banking systems, like monetary systems, have evolved in the different nations in accordance with local traditions, habits, and political institutions. In the two chapters immediately preceding the application of banking and credit principles was illustrated by the experience of the United States. Further light may be thrown upon this subject by a brief examination of the various banking organizations of some of the other leading nations.

We note first that in all the important commercial countries except Russia direct daily contact between the banking organization and the people, through the essential functions of discount and deposit, is everywhere maintained by local banks or branch banks, privately owned and privately managed. Over and above these local banks there is to be found in most of the nations a great central bank which, though not generally owned, is controlled by, or at least very closely associated with, the national government. To this central bank is usually entrusted three important services: (1) the function of note issue, which in the development of banking has quite commonly been taken away from the local banks and made a monopoly of the central bank, (2) the business of rediscount for the local commercial banks, and (3) the concentration and management of the nation's banking reserves. In addition the central bank is the banker of the government, keeps the government money, manages the public debt, and handles the government's financial business. The central bank may or may not have ordinary dealings of discount and deposit with the general public. Thus we find at the head of the English banking system the Bank of England, in France the Bank of France, central banks in Germany, Hungary, Italy, and most of the other European nations, in many countries of the Western Hemisphere, as well as in South Africa, India, China, Japan, Australia, and New Zealand.

The French banking system. This general description finds its clearest exemplification in France. The French public depends for its ordinary banking services mainly upon a few large banks located in the great cities and with branches scattered about the country. Small local banking corporations and private banks exist but are not numerous or especially flourishing. The Bank of France also deals directly with the public. Anyone may carry a deposit account with the Bank and anyone may apply for discounts or advances. In 1933 the Bank had a nation-wide system of 660 branch offices and agencies and carried more than 388,000 deposit accounts.

The Bank of France: Ownership and control. This is one of the oldest, largest, and most famous of central banks. As constituted by Napoleon at the beginning of the nineteenth century, the Bank was owned by private citizens, but although the shares were in the hands of thousands (recently around 40,000) of shareholders, only 200 principal shareholders were allowed to vote. Though it owned no shares, the government appointed the governor and the two assistant governors, and the Bank acted as the sole fiscal agent of the government. The 200 voting shareholders formed the General Assembly, from among whom fifteen regents and three censors were chosen. They, with the governor and assistant governors, formed the General Council, which controlled the Bank's policies. To be eligible as a regent a man had to own at least 85 shares, worth approximately \$50,000; the governor had to own at least 200 shares. The regents were not only rich men and leaders in finance, industry, and commerce, but their position was often in practice hereditary.

Because of this control the Bank of France has been subject in recent years to severe attack as "a feudal plutocracy," "the stronghold of resistance to popular sovereignty and the will of the State," an attack culminating in 1936 in a radical reform by the leftist government. The new governing body of the Bank is the Council of Regents, a general council of twenty-six members, including representatives of the government, the shareholders, and the Bank personnel, and also a group of members appointed by the National Economic Council to represent finance, industry (large and small),

commerce, labor, agriculture, and the consumers. The governor, who is prohibited from having any outside financial interests, and the vice-governors are still appointed by the government, but they need no longer be shareholders. Shareholders now have one vote each, and proxies may be used, but voting rights are limited to the election of three "Censors," who have advisory powers only, and two members of the Council.

This reform was obviously one of "democratization" of the Bank, without nationalization, but its effect is to make the Bank subject more than ever to political influences, since control is taken from the shareholders and lodged with a council dominated by the State.

Character of business. The Bank of France discounts domestic short-term acceptances, ordinarily requiring that they be endorsed by three or more firms, although when secured by collateral in the form of stocks or bonds only two names are required. In practice the commercial banks generally rediscount only their small bills, desiring to keep secret the names of their large customers. The Bank makes advances also against securities, including all kinds of government securities and a long list of other eligible securities. Loans against securities are called "lombard loans" and are made at "penalty" rates; *i.e.*, from one to two per cent above the discount rate. Another type of loan, very prominent in recent years, is made against deposits of gold bullion or foreign gold coins. The rate on these loans is usually half of one per cent higher than the discount rate.

The Bank possesses limited authority to operate in the open market, dealing in foreign exchange, in short-term securities for the account of foreign central banks, and in bonds of the Caisse Autonome d'Amortissement, an official agency performing the general functions of a sinking fund. Although the authority to deal in these securities, granted in 1928, has afforded the Bank a new type of credit control, the Bank has never attempted seriously to develop such open-market operations. One reason for this inaction has been the limited character of the Paris discount market, although in recent years efforts have been made to widen the market by establishing new discount houses and other facilities. The deal-

ings in foreign exchange, on the other hand, are very important in influencing credit conditions and have the same effect in France as have dealings in securities in America and elsewhere. The Bank relies heavily upon its control of discounts, and its discount rate is very influential in French finance since this rate is followed by the commercial banks and prevails in the Paris market. The Bank employs also a stabilization fund to control the foreign value of the franc and another fund to control the price of rentes.

The Bank has the exclusive right of note issue. This function is of more importance in France than in the Anglo-Saxon countries, because of the relatively undeveloped state of deposit banking and the strong reliance upon bank credit in the form of notes. The notes of the Bank of France are the principal element of the French monetary system. They are issued in connection with loans to the public, rediscount of commercial paper for the other banks, and loans to the government. They have also been given out in great volume in exchange for gold. The result is that, while no distinction is made between notes exchanged for gold and those issued in discount of commercial paper, the gold reserve back of the entire note issue has usually been very large.

The Bank may, if it pleases, redeem its notes in coin, but its legal obligation is only to convert its notes on demand into bullion when presented in such minimum amounts as the Bank and the Minister of Finance may determine. The minimum before France abandoned the gold standard in 1936 was 215,000 francs, and the gold standard was abandoned by increasing the minimum amount in which the Bank would redeem its notes in gold to 5,000,000,000 francs and subjecting gold exports to a governmental license system. The legal reserve minimum is thirty-five per cent of the circulating notes and liabilities on current account, the reserve consisting of gold bullion and gold coin.

An important fact in recent history is the remarkable accumulation of gold in the Bank of France and the leadership it exercised in the maintenance of the gold standard after 1931. The enactment of the stabilization act of 1928 was followed by an inundation of gold; other waves have occurred meantime; at the end of November, 1932, the Bank held 83,342 million gold francs; in November,

1935, it held 70,389 millions; and on June 30, 1938, its gold amounted to 55,808 millions.

While it has direct dealings with the public and is actually the largest commercial bank in the country, the Bank of France is also a bankers' bank. It holds deposits of the other banks, it rediscounts for them, and its great hoard of gold is the national reserve reservoir, to be drawn upon as needed by the other banks through rediscount or by cashing their deposits.

The following condensed balance sheet will give further evidence as to the character of the business of the Bank of France.

FINANCIAL STATEMENT OF BANK OF FRANCE, JUNE 30, 1938

(In millions of francs)

Assets:

Gold coin and bullion	55,808
Silver and copper coins	424
Postal current accounts	554
Funds available at sight abroad	27
Domestic and foreign bills and notes	6,726
Advances against securities, etc.	4,165
Bonds of the sinking fund	5,574
Loans to the government	43,449
Other assets	3,966
Total	<u>120,693</u>

Liabilities:

Capital	183
Surplus, etc.	329
Notes in circulation	102,087
Deposits	16,014
Miscellaneous	2,079
Total	<u>120,693</u>

The total demand liabilities of the bank were on this date 118,101 million francs, and its ratio of gold reserve to demand liabilities was 47.25 per cent.

British banking: Commercial banks. Deposit banking is more highly developed in England than anywhere else in the world. A comparatively small number of large joint stock banks with thousands of branches, together with a small and declining number of private banks, bring the services of discount and deposit into every corner of the country. There are now sixteen joint stock banks and four private banks, operating almost exclusively in England and

Wales (this excludes finance houses). Of the joint stock banks five — the Midland Bank, Lloyds Bank, Barclays Bank, Westminster Bank, and National Provincial Bank — tower above all others. These are known as the “Big Five,” and they control about 85 per cent of the total banking resources. The British banking system includes also the eight Scottish joint stock banks and the three North Irish joint stock banks, together with their branches and affiliates. The use of deposits with payment by check is universal, and note issue is of relatively small importance. The Bank of England now has a monopoly of note issue, the other banks which had retained their note-issue privilege under the act of 1844 having lost it, chiefly through the amalgamation process.

The Bank of England: Character of business. At the head of this banking system stands the Bank of England, founded in 1694 and with a long and illustrious history, out of which have developed peculiarities which set it apart in a class by itself. The Bank of England is in no sense a government bank, except that it is the fiscal agent of the British government. Its capital is all privately owned, and it is managed by directors chosen by the stockholders like any ordinary corporation. Though it handles all the British government’s financial business, it has the least connection with government of any of the great central banks. Yet there is always excellent coöperation between the government and the Bank. For example the Bank voluntarily turned over to the government its excess profits from the World War period.

While it deals directly with the public, through discount and deposit, the Bank of England does not wage a vigorous competition with the other joint stock banks, in whose hands is the bulk of the nation’s banking business. The Bank has few transactions of loan or discount with the joint stock banks. In volume of assets it is a small bank compared with some of these, and its loans to the market in normal times are negligible as compared with the amount loaned by the commercial banks. The Bank’s customers are bill brokers, discount houses, individuals, firms, etc., and for these it discounts at the market rate. The custom of the Bank is to announce each Thursday a “Bank rate”; that is, the minimum rate at which it stands ready to discount prime bills for others than its regular

customers. This Bank rate is higher than the market rate, and there are therefore no applications to the Bank for discount by outsiders in normal times. If for some reason, such as loss of reserves by the joint stock banks, the market tightens and the market rate rises to the Bank rate, outsiders may resort to the Bank for discounts and advances. At such times the Bank rate becomes "effective" in influencing the market situation. It is of course possible for the Bank itself to tighten the discount or security market by selling bills or securities and so absorbing funds afloat there.

Concentration of reserves. Although the Bank of England does not have continuous discount relations with the other commercial banks — standing rather as an emergency institution ready to help in crises — it is continuously a bankers' bank in regard to reserve holding. The bank holds both its own reserves and the Exchequer balances which it carries as fiscal agent and depository for the government. In addition every British bank, including those in Scotland and Northern Ireland, keeps a part of its reserve in London, either in the Bank of England or in one of the great London banks or in both. The London banks keep about half of their cash reserves on deposit with the Bank of England. Their till-money consists of course largely of Bank of England notes, and from this point of view practically the whole cash reserve consists of demand rights against the Bank of England in the form of notes and deposits. The Bank of England is thus the custodian of a goodly portion of the entire banking reserve of the United Kingdom. This concentration promotes the efficiency of the reserves, making possible a very large superstructure of credit.

Note issue. It is in the system of note issue that the peculiar internal structure of the Bank of England appears most strikingly. The system dates from the bank act of 1844, when the Bank was already 150 years old. By this act the Bank was separated into two distinct parts, called respectively the "Issue Department" and the "Banking Department." The latter conducts a regular banking business (except note issue) such as was described in the chapter preceding this. The issue of notes is in the province of the Issue Department and is based upon a principle different from that

of the Bank of France. Notes issued by the Bank of England must, with very minor exceptions, be secured either by British government obligations or by gold, and it was mainly to insure this that the Issue Department was separated from the rest of the Bank. In 1844 the government owed the Bank the sum of £11,015,100. These securities were put in the Issue Department as its assets, and it was allowed to issue notes to the same amount. Certain other joint-stock banks which had note issues outstanding at that time were permitted to retain them, but with the understanding that if any of these banks relinquished its note issue privilege by voluntary or involuntary surrender of its charter or by amalgamation with another bank or by moving its home office to London, the Bank of England would inherit a portion of the forfeited privilege. Under this proviso the volume of Bank of England notes secured by government paper increased to £19,750,000 in 1923, when the Bank attained a monopoly of note issue.

In 1928 was passed the "Currency and bank notes act," which shifted the liability of the war-time issue of government notes from the Treasury to the Bank of England and fixed the maximum amount of notes which may be issued by the Bank on security other than gold at £260,000,000. This act also gave the Treasury authority, at the request of the Bank, to reduce the fiduciary limit for any stated period or to increase it for not more than six months at a time, provided that no increase may extend beyond two years without specific sanction of Parliament. It is understood that the provisions for increase will be used if necessary (1) to meet a panic, (2) to allow for large withdrawals of gold abroad, (3) to mitigate any scramble that might arise among central banks for gold, or (4) to provide for the normal growth of population and business. In order to remove the temptation to increase the fiduciary issue for the sake of the profit involved, it is provided that the net profit made by the Bank from the securities held against the fiduciary issue shall be paid to the Treasury.

Reference to the financial statement of the Bank of England, as of June 15, 1938, on page 645, will help to make these matters clear. The first two items on the asset side of the Issue Department's statement are obligations of the British Government, the

original sum held in 1844 being recorded separately from the later additions. Of the £526,410,111 in notes outstanding on this date £326,410,111 were covered by gold coin and bullion, though not redeemable therein. The fiduciary issue at this time amounted to £200,000,000. This segregation of the note issue function in a distinct department, with specified assets to cover the notes, is peculiar to the Bank of England, no other central bank having copied the idea.

Elasticity of notes: Elasticity sacrificed. Important results emerged from this peculiar system of note issue. Disregarding for the moment the changes that came in 1931, it will be evident that the total amount of bank notes was normally quite rigid. Changes in the volume of bank notes came only as the Issue Department gave notes to the Banking Department in return for gold or received notes from the Banking Department for redemption in gold. This provided the machinery for redemption under the gold standard, but it gave no elasticity in relation to the community's need for currency. Since 1931 there has been no right of redemption of bank notes in gold. There is still however some exchange of notes and gold between the two departments of the Bank in connection with the operations of the Equalisation Account, which occasionally exchanges gold for notes or notes for gold with the Bank. This gives a certain limited elasticity to the fiduciary issue, but it leaves it still true that the notes of the Bank of England from 1844 to the present have not furnished that elastic element in the currency system which we have seen it is one of the principal purposes of bank credit to provide. The notes of the Bank of England have been given perfect security but at the expense of elasticity.

How Great Britain secures elasticity. If it be asked how Great Britain has managed with a currency system so seriously defective in this respect, the answer is twofold. In the first place, as has been seen, the British have come to rely upon bank deposits to an extraordinary degree, and the deposit system of Great Britain has well-nigh perfect elasticity. Secondly the gold element in the British monetary system has been elastic to a degree found nowhere else in the world. London has long occupied a peculiar position as the financial and monetary centre of the world. Here is the world's

STATEMENT OF THE BANK OF ENGLAND, JUNE 15, 1938

<i>Issue department</i>		
		£
Liabilities —	Assets —	
Notes issued :	Government debt	11,015,100
In circulation	Other government securities	188,933,755
In banking department	Other securities	43,781
	Silver coin	<u>7,364</u>
	Amount of fiduciary issue	200,000,000
	Gold coin and bullion	<u>326,410,111</u>
Total	Total	<u><u>526,410,111</u></u>
<i>Banking department</i>		
	Assets —	£
Liabilities —	Government securities	114,401,164
Proprietors' capital	Other securities	25,773,844
Rest	Discounts and advances	(5,680,689)
Public deposits	Securities	(20,093,155)
Other deposits	Notes	40,672,675
Bankers	Gold and silver coin	<u>856,646</u>
Other accounts	Total	<u><u>181,704,327</u></u>
Total	Ratio of reserve to liabilities	25.3%

greatest market for notes, drafts, bills of exchange, and all forms of commercial paper. A rise in the rate of interest in London causes a flow of capital from the ends of the world seeking the favorable opportunity for investment, while a fall in the interest rate checks the inflow or even starts a movement in the opposite direction. The current interest rates in London are strongly influenced, if not controlled, by the rate of discount of the Bank of England. Thus the Bank of England has until recently held the lever by which it controlled the international flow of gold, supplementing the natural tendency of gold to move to that country which especially needed it, as shown by a relatively low price level, and *vice versa*. Since 1931 control is more largely effected through gold transactions than through changes in the Bank rate.

Extreme procedure in time of crisis. There have been, it is true, certain occasions when the British credit system, functioning well enough in ordinary times, has broken down. In time of financial crisis there is need of great and sudden expansion of bank credit, or, more correctly, there is need of expansion at some point to make good the panicky contraction of the regular sources of credit. Under such circumstances reliance was in other countries upon the notes of the central bank, but the Bank of England, with its rigid system of note issue, has on several occasions been unable to meet this demand. It was then found necessary for Parliament to "suspend the bank act" by permitting an additional "uncovered" issue of notes; that is, of notes not secured by gold. This rather curious and clumsy device has produced the desired result of promising sufficient credit and allaying the panic. In 1928, as noted above, the method of suspending the bank act was made less awkward, as Parliament gave the Treasury a continuing authority to authorize the Bank to increase the fiduciary issue by such amount as the Treasury might think proper.

The Banking Department. A few words only will suffice to explain the second part of the Bank of England's statement as presented on page 645. The liabilities of the Banking Department include the capital stock ("proprietor's capital"), surplus ("rest"), deposits of the government ("public deposits"), and other deposits, a large part of which represents the Bank's obligation to the other

banks. On the asset side we find government securities, "other securities" (which here means loans and discounts and all securities except government obligations), notes, and gold and silver coin. The notes are the notes of the Bank itself, which, in the statement of the Banking Department, are assets; that is, they are claims of this Department against the Issue Department. Until Great Britain's departure from the gold standard in 1931, the Banking Department could get notes at any time from the Issue Department in exchange for gold, and it could likewise always demand gold from the Issue Department to redeem bank notes which it might hold.

The reserve. By combining the two parts of the statement it will be seen that the Bank on this date had a net note liability of 485 million pounds, deposit liabilities of 164 millions, and a gold reserve of 327 millions, or practically 50 per cent of its liabilities to other than its stockholders. This was about the percentage which generally prevailed before the World War, testifying to the Bank's position as the custodian of the nation's reserves and its preparedness to meet the demands of the other banks, the government, and the British public. During the war the reserve ratio fell as low as 23 per cent, through loss of gold and expansion of credit. These comparisons of reserve position are however not strictly valid, for several reasons: gold no longer circulates in Great Britain but is all concentrated in the central bank reserve or the Treasury funds, the amount reported by the Bank of England does not include the gold in the equalization account, the Bank's notes are no longer redeemable on demand, and the gold reserve can be protected by bidding up the price of gold in the open market.

The Canadian banking system: The chartered banks and their resources. As was stated at the beginning of this chapter, a great central bank more or less closely related to the government stands at the head of the banking system in the majority of the leading nations of the world. There is however one great nation which maintained for a generation and more a highly efficient banking system without the aid of a central bank, although a central bank was finally established in 1935. A brief study of the Canadian banking system will fully reward our time and attention. Banking

is here in the hands of a few large "chartered" banks, each of which maintains a great number of branches scattered all over the dominion, thus bringing the facilities of one or more banks to the doors of practically every inhabitant. The business of discount and deposit is conducted in the usual way.

In 1890 the Canadian banking system embraced 38 banks. These were mostly large institutions, their average paid-up capital being over \$1,500,000, and the law provided that no new bank should be chartered with a capital of less than \$500,000, of which at least half must be paid up. Since then the number of banks has steadily diminished, in part through failures and liquidations, but mainly because of consolidations. In 1901, there were 34 banks, ten years later (1911) the number was 29, and the more rapid decline of the next dozen years brought the number down to 10, where it stands at present (July 31, 1938). At the same time the size of the banks increased, as well as the number of their branches. The average paid-up capital, which was over \$1,500,000 in 1890, had become practically \$2,000,000 in 1901, nearly \$3,500,000 in 1911, and on December 31, 1934, was \$14,450,000, on which date the average surplus was \$16,200,000. The number of branches was 426 in 1890, about 1,650 in 1907, 1,841 in 1911, and 4,147 in 1928. During the depression the banks have closed many unprofitable branches, the net reduction in number from the 1928 figure being 634 by the end of 1937. On May 31, 1938, the total paid-up capital and surplus of the ten existing banks was \$279,250,000, and their total resources were \$3,331,523,000.

Notes. As is stated later, the Bank Act of 1934 planned to end the note issue of the chartered banks. Before that law took effect all of the chartered banks issued notes, secured by their general assets in the normal manner, but having a prior lien ahead of other liabilities. The notes were further secured by a "safety fund," held and administered by the government, to which each bank contributed an amount of money equal to five per cent of its note issue. In case of failure of any bank to meet its note liabilities, the notes were paid out of this fund, which then had to be brought back to five per cent of the total note circulation, either by sale of the assets of the failed bank or if necessary by further contributions

from the other banks. Canada thus obtained well-nigh perfect safety for her bank note currency, without the earmarking of government bonds or any other particular assets of the banks as security.

The Canadian bank notes were also perfectly elastic. They were issued through the branches all over the country in connection with the discount of commercial paper. Each bank redeemed on demand its own notes and also accepted the notes of the other banks, which were immediately presented to the bank of issue for redemption through the clearing houses in the large cities and through the nearest branches in the rural sections, in exactly the manner with which we have become familiar in the case of checks. The note circulation thus expanded and contracted in response to business needs with the same facility as deposits and bore the same relation to bank reserves. This emphasis on the note issues of the Canadian banks must not be interpreted to mean that note issue is more important than deposit banking; for instance, on June 30, 1938, of the 3,339 millions of dollars of total liabilities of the banks 2,459 millions were deposits and 99 millions were bank notes.

Record and judgment of Canadian banking. The record of the Canadian banking system for safety and efficiency has been excellent. Failures have been few, with no losses whatever to note-holders and only moderate losses to other creditors. Only one bank has failed since 1914. This is in wide contrast with the failure record in the United States, which resembles Canada in geographical extent and industries. The one failure was in 1923 and was due to crookedness of officers rather than to weakness of system. The freedom of Canada from bank failures cannot be explained on grounds of ultra-conservatism and inadequate provision for banking services. Branch banking has enabled the banks to extend to all parts of the country, both frontier and otherwise, the varied services of the great metropolitan institutions. The paucity of failures is due chiefly to the following causes: the branch bank structure of the system, making runs difficult and less likely to occur, the large size of banks and the high order of administrative officials, the illegality of loans by the banks on real estate, the incorporation of the banks by a single federal authority, the wide distribution of

the banks, affording automatic diversification of loans as regards locality and industry, the economy of operating branches instead of completely staffed and equipped unit banks, the support which the parent bank can give to a local bank suffering from emergency or cyclical difficulties, and the annual audit of the banks by public accountants engaged by the shareholders.

The accompanying statement shows the condition of the ten chartered banks as a whole on June 30, 1938:

COMBINED STATEMENT OF CANADIAN CHARTERED BANKS
June 30, 1938
(In millions of dollars)

<i>Assets</i>		<i>Liabilities</i>	
Cash in Canada (B. of C. notes and deposits)	247	Notes	99
Securities:		Deposits payable in Canada:	
Dominion and provincial short	437	Demand	838
Dominion and provincial long .	719	Time	1,621
Other	306	Total deposits . . .	2,459*
Total securities	1,462	All other items	781
Loans:			
In Canada:			
Call	69		
Current public	786		
Current other	135		
Loans abroad and net due from foreign banks	119		
Other assets	520		
Total assets	3,339	Total liabilities . .	3,339

* Excludes interbank deposits.

The Bank of Canada. The Canadian bank law has been revised every ten years. Because of the depression, the 1933 revision was delayed, and a sweeping inquiry was made by the "Macmillan Commission" into the whole financial system. The commission reported that "the Canadian banks give admirable evidence of security, efficiency, and convenience," but it recommended among other things the establishment of a central bank. The want of a central bank at which the chartered banks could rediscount and borrow had been felt during the World War, and the emergency arrangement of loans from the government had proved unsatis-

factory. The commission now recommended the establishment of a real central bank to take over this function, to promote the free operation of the gold standard, to control and defend the value of the Canadian dollar abroad, to coöperate with other foreign central banks, to influence the degree of business activity, and to serve and advise the government.

The Bank of Canada (established in pursuance of the Bank Act of 1934) had originally a capital of \$5,000,000 in 100,000 shares valued at \$50.00 each. These were offered to the public and taken up at the par value. By law the shares were to be owned only by British subjects, ordinarily resident in Canada, and not by any chartered bank. Dividends on these shares were limited to $4\frac{1}{2}$ per cent a year, and any excess profits were to become the property of the Government. The Bank was to be operated by a governor and deputy governor appointed by the Government in the first instance but later by the directors, of whom there were seven, elected by the shareholders.

Change of political control of the Government, effected by the 1935 elections, led to amendment of the bank act, giving to the Government ownership of a majority of the shares. This was done by the issue of new capital of \$5,100,000, in 102,000 shares, to be held by the Minister of Finance on behalf of the Government. The amendment provided also for the appointment by the Government of six directors with two votes each, thus insuring to the Government a majority over the seven shareholders' directors having one vote each. At the same time provision was made for a gradual reduction of the directorate to seven.

Finally in 1938 legislation was enacted providing for the complete nationalization of the Bank. The capitalization was reduced to the original figure of \$5,000,000, all owned by the Government, the Government redeeming the stock held by private persons at \$59.20 a share.

The Bank is designed to supplement and regulate the banking system in Canada but not to compete with the chartered banks. It may not accept deposits from or make discounts or loans to anyone except the chartered banks, the Dominion and Provincial Governments, and banks incorporated under the Quebec Savings Bank

Act. The Bank will gradually acquire a monopoly of note issue; the issue of notes of the chartered banks is restricted, and their notes will ultimately be displaced. The chartered banks are required to keep a minimum reserve of five per cent of their deposits within Canada, the reserves to consist of balances with the Bank of Canada and notes of this Bank.

The Bank of Canada assumed the Government's liability for the Dominion notes outstanding and received from the Treasury gold, silver, and securities to cover them. The debts of the chartered banks to the Treasury are to be liquidated. All the gold held by the chartered banks was transferred to the Bank of Canada at \$20.67 an ounce, and any profits from future devaluation of the Canadian dollar are to accrue to the Government. The Bank is to keep a 25 per cent gold reserve against its notes and demand liabilities. The Bank is authorized to hold a limited amount of silver (valued at the market price) and first-class foreign exchange. No specific or segregated assets are required as security for notes. The Bank has no separate issue department, like the Bank of England, nor does it resemble the American Federal Reserve in having regional banks and member-bank ownership.

STATEMENT OF BANK OF CANADA
(November 2, 1938, in millions of dollars)

<i>Assets</i>	<i>Liabilities</i>
Reserve:	Capital paid up 5.0
Gold coin and bullion . . . 181.0	Rest Fund 1.3
Silver bullion 0.0	Notes in circulation 177.0
Foreign currencies 36.6	Deposits:
Total reserve 217.6	Dominion government . . . 27.8
Subsidiary coin2	Chartered banks 221.5
Securities:	Other 1.4
Dominion and provincial:	Total deposits 250.7
Short 165.5	All other liabilities 5.1
Long 44.8	
Total securities 210.3	
Advances to:	
Chartered and savings banks . 5.0	
Bank premises 1.7	
All other assets 4.4	
Total assets 439.2	Total liabilities 439.2

The statement on the opposite page, as of November 2, 1938, will show the character of the Bank's assets and liabilities.

It will be observed that the Bank's reserves consist of gold, silver, and foreign currencies, and amount to 50.9 per cent of the sum of its notes and deposits due to the chartered banks, the government, and the public.

XX

XXVIII

MONEY AND PRICES

The value of money. As we define and use the terms "value" and "price" in this book, the value of anything is the quantity of some other thing that would be given in exchange for it, and the price of anything is the quantity of money that would be given in exchange for one unit of it.¹ In practice it is customary to express values as well as prices in terms of money. But what then is the value of money itself?

Although we are completely habituated to measuring the values of other things in terms of money rather than the other way around, there is no reason why money should not conform to the same concept of value as applies to anything else. The value of any commodity may be expressed in terms of any other commodity. So may the value of money be expressed. If the value of four pounds of coffee is one dollar, then the value of one dollar is four pounds of that particular brand of coffee. There is a value of money in terms of wheat, a value of money in terms of leather, another in terms of shoes, another in terms of theatrical entertainment, and so on. But to express values in terms of other things indiscriminately would for most purposes be clumsy and confusing, and so by common consent we have come to express all values (except the value of money itself) in terms of money. In like manner it would be unprofitable to express the value of money, now in terms of one commodity, now in terms of another.

We obviously cannot escape this difficulty by expressing the value of money in terms of itself. Moreover the concept of the value of money, to be useful, must not be its relation to some one commodity. What we are interested in is the relation of money to all other commodities and services, to economic goods in general. When we say money has depreciated in value, we mean that one

¹ See Chapter I.

unit or any other given amount of it will not buy so much; and we do not have in mind one particular commodity; we think of it as less valuable with respect to anything and everything that could be bought. So when money is of high value we mean that it will buy much, or conversely that much of other things must be offered to obtain any given quantity of money. This is evidently nothing else than what people mean by the "purchasing power" of money. We thus arrive at a satisfactory working definition of the value of money and one that agrees closely with the popular concept. *The value of money is the quantity of other things in general that will be exchanged for one unit of money; i.e., its purchasing power.*¹

It will of course be clear that other definitions of the value of money, lying between the value expressed in terms of a particular commodity and the value expressed in terms of all other things in general, are possible. We may thus express the value of money in terms of a limited list of selected commodities and services, such as those that are commonly consumed by most people, or those that are of special interest to particular groups or classes of the people. For the practical purposes of measuring the value of money, as we shall see, it is always necessary to select certain goods; to use all goods and services that exist would be manifestly impossible. We may thus have any number of special definitions for specific purposes. The most useful concept of the value of money however is that which is expressed in the most general terms, as defined above. Other concepts may best be regarded as approximations to or special cases under the generalized definition.

Value of money not constant. The universal custom of measuring all values in money is responsible for the very common and persistent notion that the value of the monetary unit is a fixed thing and that all price changes are due to changes in the values of the priced commodities. There is scarcely any economic truth whose

¹ It will be noted by the careful reader that, as we have defined the term, value may relate to any quantity of the good in question (not merely to a unit). When we speak of the value of anything, as of land, we must specify the quantity, as a ten-acre farm, or a city lot. When referring specifically to the value of a unit of anything, we generally use the term price. In referring to the value of a unit of money we cannot say price, since we have chosen to express prices in terms of money. Hence, unless otherwise stated, we shall use the phrase "value of money" to mean the value of one unit of money.

popular acceptance is more difficult to compel than the fact that the value of money is not constant. Yet a moment's consideration of the definition of money should convince anyone of the truth of this principle. Value is always a relationship between two things. If one bushel of wheat exchanged for two bushels of corn yesterday and for three bushels of corn today, we may say that the value of a bushel of wheat has *risen*, from two bushels of corn to three bushels, assuming thus that the value of corn is a constant. But we could just as correctly assume the value of wheat constant and say that the value of a bushel of corn has *fallen*, from one-half bushel of wheat to one-third bushel. Without knowledge of the facts of the particular case, the change in relation may have been due to a change in either corn or wheat alone or to changes in both of them. If two cars happen to be at the same level in an elevator shaft and later it is found that car A is above car B, the change in relation may have been caused by (1) a movement upward of A while B stood still, (2) a downward movement of B while A stood still, (3) an upward movement of A and a downward movement of B, (4) upward movements of both, A moving farther than B, or (5) downward movements of both, B moving farther than A. And, to complete the case, both cars might move the same distance either up or down, leaving their relative positions unchanged. So it is of the value relation between any two commodities.

This may be further clarified by contrast with measurement in a unit that is constant. The foot is known to be to all intents and purposes a constant measure of length. Suppose we have determined by measurement that a certain room is exactly twenty feet long. If now, upon returning at a later date, we should find that the same room had come to be twenty-four feet long, we would naturally and rightly infer that it was the room (not the foot) which had changed — that a partition had been knocked out during our absence, or an extension added on. Suppose however that the foot (like the dollar) were a variable unit. This would complicate our problem. We should now not know whether the room had been lengthened during the interval, or the foot shortened, or whether both had changed. For example, the foot might have been reduced to the equivalent of ten inches of the old kind, in which case the

explanation of the room's new length in feet would not involve reference to any change whatever in the room.

Measuring the value of money. But how is the value of money, or its purchasing power, to be measured? The answer must be found in that which expresses at once the relation of money to all other commodities and services. This is price, not the price of any particular thing, but a composite price of things in general, or what we have called the "general price level." It is change in the general price level which indicates change in the value of money. A change in a particular price, while other prices remain substantially the same, is presumably due to changes affecting the particular commodity. But when the general price level rises or falls, we may safely assume that the cause is a change in the value of money. High prices and cheap money mean the same thing. Low prices and dear money are synonymous. A change in the general price level means an opposite change in the value of money.

It follows that to measure and record the value of money we must have some way of expressing the general price level. For this purpose there have been developed *index numbers*, which are capable of reflecting with a high degree of accuracy changes in the general price level or the purchasing power of money. Before proceeding further in our study of the value of money and its relation to prices, we must acquire a working knowledge of the technique of index numbers.

Index numbers: Price relatives and index numbers. Prices do not move in unison. Some rise, while others fall, and still others remain stationary. A complete picture of all price changes in any given period would show a kaleidoscopic confusion of thousands of separate and apparently unrelated prices rising or falling or standing still. Yet there is such a thing as a general price movement, just as there may be a movement of a whole swarm of bees while the individual bees are dashing about in all directions in the utmost apparent confusion and without seeming relation to each other.

For a simple example, let us consider the average wholesale prices of three commodities, bituminous coal, wheat, and copper, in the three years 1913, 1914, and 1915. Here are the facts:

PRICES			
	1913	1914	1915
Coal	\$1.27 per ton	\$1.17 per ton	\$1.04 per ton
Wheat	.91 per bu.	1.04 per bu.	1.34 per bu.
Copper	.15 per lb.	.13 per lb.	.17 per lb.

Evidently the price of coal declined during this period, the price of wheat rose, and the price of copper fell from 1913 to 1914 and rose in the next year. To measure these separate price movements we may call one year, say 1913, the base year and express each price of each commodity as a percentage of the price of that commodity in the base year. For example the prices of coal were: in 1913, \$1.27; in 1914, \$1.17; in 1915, \$1.04. Calculating the percentages we have: $\frac{1.27}{1.27} = 1$, or 100%; $\frac{1.17}{1.27} = .92$, or 92%; $\frac{1.04}{1.27} = .82$, or 82%. If then we let 100 stand for the price of coal in 1913, the price in 1914 relative to it will be 92, and the 1915 price, 82. Such numbers are called *price relatives*. By similar calculations we find the price relatives for wheat and copper with respect to their prices in 1913, and the whole result is shown in the following table:

PRICE RELATIVES			
	1913	1914	1915
Coal	100	92	82
Wheat	100	114	147
Copper	100	87	113

Such a table might be continued to include similar price relatives for all other important commodities, giving thus a record of the price changes of each separate commodity. But what we now desire is to know how prices in general have changed; we are seeking a record of changes in the general price level. For this there is required a number which shall express for each year the general or average price relative, evidently the average of the separate price relatives for each year. Such a number is called an *index number of prices* and may be defined thus: *An index number of prices is a number which expresses the general price level for any given year or other period relative to the general price level of some particular period which*

*is taken as the base period and whose general price level is commonly represented by 100.*¹

The average, simple and weighted. There are many kinds of averages, and the particular average chosen will determine the character of the index number. For example we might employ the simple arithmetic mean of the separate price relatives of each year. In our example the average for 1913 is of course 100. For 1914 the simple arithmetic average is $\frac{92 + 114 + 87}{3}$, or 98. For 1915, it is

114. Index numbers have been constructed on this principle, but they have a serious defect. The several commodities which enter into trade are not of equal importance, and an index number which takes no account of such differences is not an accurate record of the changes in the general price level or the purchasing power of money. The value of money is much more affected, for instance, by a change in the price of wheat than by a change in the price of clover seed.

To avoid this defect of the simple arithmetic mean it is necessary to use a weighted average which takes proper account of the relative importance of the several commodities. For example suppose that in a given year the price of corn increased four per cent, while the price of butter increased ten per cent, and that corn was twice as important as butter. The simple arithmetic average would show a combined increase of seven per cent; $\frac{4 + 10}{2} = 7$. But this is not a

true picture, since no account is taken of the greater importance of corn. The average is corrected by putting the figure for corn in twice; thus $\frac{4 + 4 + 10}{3} = 6$. This is a weighted average of the

percentage increases. Weighting may be accomplished in various ways, the most satisfactory method being to weight the price movement of each commodity according to the total value of that commodity exchanged in a given year. Thus if twice as much money is

¹ The term "general" in this definition does not imply that the prices of all commodities are included in any given index number. The meaning is that the index number represents the result of averaging price relatives for all commodities which have been considered. It is only in this sense that any index number shows a "general price level." Actually it is possible to construct index numbers which are capable of serving almost as well as though they were in fact based upon all commodities.

expended for commodity A as for commodity B in the selected year, this is taken to indicate that A is twice as important as B and should be given double the weight of B in the index number. The simplest way to accomplish this is to calculate first the total value of all the goods exchanged in a given year and then to compare with this the value of exactly the same quantities of these goods reckoned at the prices of each of the other years. To make the method clear, let us return to our previous example. Here are the facts for the first year :

YEAR 1913			
	<i>Prices in 1913</i>	<i>Quantities exchanged in 1913</i>	<i>Values at 1913 prices</i>
Coal	\$1.27	447 million tons	568 million dollars
Wheat	.91	555 million bu.	505 million dollars
Copper	.15	812 million lbs.	122 million dollars
Total value			1,195 million dollars

It cost \$1,195,000,000 to buy these quantities of these commodities in 1913. If the same quantities of these commodities cost more or less in 1914, it must have been because of changes in their prices, and if we could determine what the cost was in 1914 we should have a very good measure of the average movement of these prices. This may be readily calculated, as follows :

YEAR 1914			
	<i>Prices in 1914</i>	<i>Quantities exchanged in 1913</i>	<i>Values at 1914 prices</i>
Coal	\$1.17	447 million tons	523 million dollars
Wheat	1.04	555 million bu.	577 million dollars
Copper	.13	812 million lbs.	106 million dollars
Total value			1,206 million dollars

Evidently commodities which cost \$1,195,000,000 in 1913 were worth \$1,206,000,000 in 1914. There must have been a small rise in the average price level of these three commodities.

For the year 1915 similar results are obtained, as follows :

YEAR 1915			
	<i>Prices in 1915</i>	<i>Quantities exchanged in 1913</i>	<i>Values at 1915 prices</i>
Coal	\$1.04	447 million tons	465 million dollars
Wheat	1.34	555 million bu.	744 million dollars
Copper	.17	812 million lbs.	138 million dollars
Total value			1,347 million dollars

Again there has evidently been a rise in the average price level.

An index number of the aggregative type. The final step in deriving a real index number of the prices of these three commodities is to express the total value for each year as a percentage of the total value in 1913. Thus: $\frac{1,195}{1,195} = 1.00$, or 100%; $\frac{1,206}{1,195} = 1.01$, or 101%; $\frac{1,347}{1,195} = 1.13$, or 113%. These percentages are the index numbers and the series may be thus expressed:

<i>Year</i>	<i>Index number of prices</i>
1913	100
1914	101
1915	113

This represents an index number of the *aggregative* type. The year 1913 is the base period, its index number being 100. The weighting is according to the values exchanged in 1913. Though the same in our example, the year chosen for the weighting need not necessarily be the base year.

Practical index numbers. The example used in the foregoing explanation was purposely made simple by employing only three commodities. A practical index number must take account of a large number of articles and may be calculated for any number of years, but no new principle is thereby involved. It must also be noted that the weighted arithmetic mean, which we have used, is by no means the only possible average. The geometric mean has certain advantages, and the theoretically ideal formula is more complicated than either of these. It is not necessary for us to follow this path further. Fairly satisfactory results are obtained from each of several formulas. The aggregative type of index number which we have studied serves well to bring out the essential principles of index numbers in general, and it is also one of the most important and on the whole probably the best of the types in actual practical use.¹

¹ A standard may fall short of perfection, and yet serve fairly well until refinements make it more exact. Thus Professor Fisher points out that the yardstick itself "has gone through many stages — for instance. (1) the girth of the chief; (2) the length of the arm of King Henry I; (3) the length of a bar of iron in the

Two of the most useful series of index numbers and the ones most generally used in the United States are those published by the United States Bureau of Labor Statistics. The first of these was originally published in 1902, since which time it has undergone various changes from an unweighted mean of price relatives to its present form, which is an example of the aggregative type described above. It shows the index number of wholesale prices for each year from 1890 to 1926 and for each month from January, 1913, through 1926. The number of commodities used has varied from time to time. For the first year, 1890, there were 192 commodities or separate series of quotations; in 1914 the number had increased to 297; in 1926 it was 404. The base year is 1913, and the weighting, utilizing the latest Census reports of quantities exchanged, is made in accordance with the values exchanged in the year 1919. There is now also a revised series, which uses 1926 as the base year and is weighted in accordance with the values exchanged in the three years, 1923, 1924, and 1925. This series employs 550 commodities or price series. Annual and monthly numbers are given. The table on page 664 gives these two annual series. The figures in the second column under the 1926 base are the reciprocals of the corresponding index numbers, thus showing the purchasing power or value of the dollar in each year relative to its value in 1926 (which is called 100).

As an example of the monthly series of index numbers of the United States Bureau of Labor Statistics, the figures on page 665 are given for 1937 and 1938 (new series), to which we have again added the reciprocals showing the monthly changes in the purchasing power of the dollar.

For many purposes the most useful way to present an index number is by means of a graph or curve. The Bureau of Labor Statistics' annual index numbers of wholesale prices are thus shown in Figure 37.

Tower of London; (4) a certain fraction of a quadrant between the earth's pole and equator; (5) the distance between the centers of two scratches on two gold plugs in a bar of a special metal called 'invar,' the bar being kept in a glass case in a vault at a temperature as nearly constant as possible. There is now talk of using a wave length of light set at a certain point in the spectrum." Irving Fisher, *100 Per Cent Money*, 1935, p. 88.

There are many other practical index numbers, computed and published by various agencies in America and Europe, using various formulas and relating to the prices of selected groups of commodities in the respective countries. The United States Bureau of Labor Statistics publishes periodically index numbers for the leading foreign countries. An interesting weekly index number of American wholesale prices is computed and published in the newspapers by Professor Irving Fisher.

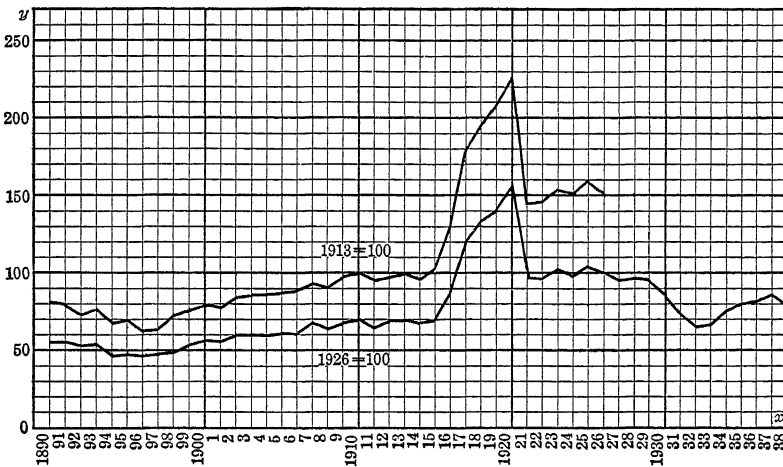


FIG. 37. Wholesale Prices in the United States

The problem of the price level: Introduction. We must now inquire into the forces that determine the value of money and the general price level. Complete exploration of this subject would lead us into one of the most complicated subjects with which economics has to deal; namely, the trade cycle or business cycle. The investigation of that subject will be undertaken at a later point in this book. In the meantime however we may dispose of certain technical factors of the problem which are directly associated with the subject of money. The general price level is immediately controlled by these factors; namely, the volume of trade, the quantity of currency, and the velocity of circulation of currency. These factors are all subject to more or less continuous and independent changes.

WHOLESALE PRICES AND VALUE OF MONEY IN THE UNITED STATES
(1890-1938)

<i>Year</i>	<i>1913 Base</i>	<i>1926 Base</i>	
	INDEX NUMBER	INDEX NUMBER	PURCHASING POWER OF THE DOLLAR
1890	81	56.2	177.9
1891	80	55.8	179.2
1892	75	52.2	191.6
1893	77	53.4	187.3
1894	69	47.9	208.7
1895	70	48.8	204.9
1896	67	46.5	215.1
1897	67	46.6	214.6
1898	70	48.5	206.2
1899	75	52.2	191.6
1900	81	56.1	178.3
1901	79	55.3	180.8
1902	84	58.9	169.8
1903	86	59.6	167.8
1904	86	59.7	167.5
1905	86	60.1	166.4
1906	89	61.8	161.8
1907	94	65.2	153.4
1908	90	62.9	159.0
1909	97	67.6	147.9
1910	101	70.4	142.0
1911	93	64.9	154.1
1912	99	69.1	144.7
1913	100	69.8	143.3
1914	98	68.1	146.8
1915	101	69.5	143.9
1916	127	85.5	117.0
1917	177	117.5	85.1
1918	194	131.3	76.2
1919	206	138.6	72.2
1920	226	154.4	64.8
1921	147	97.6	102.5
1922	149	96.7	103.4
1923	154	100.6	99.4
1924	150	98.1	101.9
1925	159	103.5	96.6
1926	151	100.0	100.0
1927		95.4	104.8
1928		97.7	102.4
1929		95.3	104.9
1930		86.4	115.7
1931		73.0	137.0
1932		64.8	154.3
1933		65.9	151.7
1934		74.9	133.5
1935		80.0	125.0
1936		80.8	123.5
1937		86.3	115.8
1938		78.6	127.2

WHOLESALE PRICES AND VALUE OF MONEY IN THE UNITED STATES,
BY MONTHS, 1937-1938

Month	1937		1938	
	INDEX NUMBER OF PRICES	PURCHASING POWER OF THE DOLLAR	INDEX NUMBER OF PRICES	PURCHASING POWER OF THE DOLLAR
January	85.9	116.4	80.9	123.6
February	86.3	115.9	79.8	125.3
March	87.8	113.9	79.7	125.5
April	88.0	113.6	78.7	127.1
May	87.4	114.4	78.1	128.0
June	87.2	114.7	78.3	127.7
July	87.9	113.7	78.8	126.9
August	87.5	114.3	78.1	128.0
September	87.4	114.4	78.3	127.7
October	85.4	117.1	77.6	128.9
November	83.3	120.0	77.5	129.0
December	81.7	122.4	77.0	129.9

Volume of trade. The chief service performed by money is the making of exchanges. From the analogy of demand and supply and the value of commodities in general, it should be evident that the value of money is dependent in part upon the volume of exchanges, since it is this that measures the need of money. The more money work there is to be done, as measured by the volume of exchanges, the greater is the need of money, and the greater will be the value of a unit of it; and *vice versa*. As our first step we conclude that the value of money tends to vary directly with the amount of trade.

Quantity of currency. It is by now a commonplace that the value of a unit of anything (*i.e.*, its price) tends to vary inversely with the quantity available for use, the result working out through supply. In a season when there is an unusually large wheat crop we expect, other things being equal, to see the price of wheat comparatively low. A crop failure generally leads to a high price of wheat. So it is with money; the more dollars there are to do a given amount of money work, the less is the value of one dollar; and *vice versa*. In other words, the value of a unit of money tends to vary inversely with its quantity.

All of the money work however is not performed by money, and in determining the value of money account must be taken of the

other forms of currency. The only important one is bank deposits subject to check. The quantity of bank deposits has the same relation to the value of money as the quantity of money itself. The conclusion of the previous paragraph should therefore be modified to read: the value of a unit of money tends to vary inversely with the quantity of currency.

However the quantity of money and the quantity of bank deposits are not wholly independent magnitudes. As we have learned, the amount of deposits which the banks have outstanding is governed in part by the amount of their cash reserves. When the ratio of reserves to deposits is high, the banks are likely to encourage loans and so increase their deposits. When the reserve ratio is low, loans are curtailed and deposits checked. There tends thus to be at least a rough relation between reserves and deposits. And since the division of the total stock of money between bank reserves and other uses depends upon habits of the people which have a certain degree of persistence, we can discern a relation between the total quantity of money and the deposits of the banks. Moreover causes which would lead to increases or decreases in bank deposits, such as business activity or depression, tend also to have the same effect upon the volume of money, principally through changes in the amount of bank notes, and this contributes also to maintaining the relation between money and deposits.

While not neglecting this relation, we must however avoid the opposite error of overlooking the possibility of considerable independent variations in the respective quantities of money and bank deposits. And we must therefore include both of these magnitudes in our analysis of the quantity of currency as a factor in determining the value of money.

Obviously the value of money can be affected only by such currency as is actually performing the monetary function of a medium of exchange. When, as sometimes occurs, the people are hoarding unusual quantities of money, such money is for the time being withdrawn from use and ceases to be a part of the quantity of currency that affects the value of money. Of course there is always a considerable amount of money in pocketbooks, in the homes of the people, and elsewhere held for later use. The difference between

quick spending and hoarding is after all one of degree. So long as the money so held is the normal amount as dictated by custom, such money should be considered as actually employed as part of the medium of exchange. Only an abnormal amount withdrawn for hoarding may properly be regarded as outside the sphere of influence upon the value of money. The variable influence of money normally withdrawn, in varying degree, from active circulation is properly taken into account in connection with the velocity of circulation of money, to which we shall give attention in the next section.

In like manner, the value of money is affected by bank deposits only so far as they are available for making current purchases. This is why savings bank deposits and time deposits generally are not included.

That the amount of currency is a variable quantity scarcely needs demonstration. Changes in one or more of the elements of the currency system are matters of frequent — in some cases daily — record. We do not have to know the precise figures for the import of precious metals from the New World into the Old in the sixteenth century to realize that this was a period when the countries of Europe were being flooded with new money. Coming closer home, we may note that the volume of individual bank deposits subject to check is estimated to have been 22 billion dollars in 1926, 23 billions in 1929, and 15 billions in 1933.

Velocity of circulation. Unlike most other commodities, money is not consumed when used. A coin or note, after facilitating one exchange, is ready with unimpaired efficiency to perform another exchange.¹ The efficiency of a given quantity of money to make exchanges is thus materially affected by its rapidity of turnover. A single dollar, if used for the purposes of exchange only once in a given period of time, must count as one in its effect on the prices of the goods and services exchanged within that period; but if a single dollar is used a dozen times within that same period, then it does the work of twelve and must count as twelve. One dollar used or “turned over” twelve times comes to the same thing as twelve dollars used once apiece. The number of times on the average that

¹ The fact that paper money wears out is not of significance, since the worn-out notes are always automatically replaced by new ones.

a unit of money changes hands during a certain period of time, say a year, is called the "velocity of circulation" of money. The amount of money work that can be done in a given time is the product of the quantity of money multiplied by its velocity of circulation, and changes in velocity have the same effect on the value of money as proportional changes in its quantity. This principle is of course equally true of other forms of currency, in particular of bank deposits. It has been estimated that money changes hands on the average in the United States about 25 times a year. The velocity of circulation of bank deposits is estimated to be somewhere between 50 and 60 times a year. We conclude therefore that the value of money tends to vary inversely with the velocity of its circulation and the velocity of circulation of bank deposits.

The quantity theory of money: Algebraic statement. Summarizing these separate conclusions, we find that the value of money tends to vary inversely with the quantity of currency and its velocity of circulation and directly with the volume of trade. Since it is the general price level which measures the value of money, varying inversely with it, our conclusion may equally well be stated thus: *the general level of prices tends to vary directly with the quantity of currency and its velocity of circulation and inversely with the volume of trade.* This principle has come to be known as the "quantity theory of money."

The reader may possibly be helped to a clearer understanding of these relations if they are put in the form of an algebraic equation. Let M stand for the average quantity of money in circulation during the year, and let M' denote the average amount of bank deposits subject to check and all other forms of currency. (The other forms are relatively insignificant and may be disregarded.) Let V and V' denote the velocity of circulation per year of money and of other forms of currency respectively. Let T stand for the volume of trade; that is, the total number of units of goods and services exchanged for currency during the year; and let P represent the general price level; that is, the average price paid for all the units of goods and services so exchanged. We then have the following equation:

$$MV + M'V' = PT$$

The quantity of money multiplied by its velocity of circulation is evidently the total amount of money paid in making exchanges during the year. Likewise the quantity of deposits multiplied by their velocity of circulation gives the total amounts of payments by transfer of deposits.¹ The sum of these two products is the total amount of payments, in all kinds of currency. The volume of trade multiplied by the average price is obviously the total value of all things exchanged for currency. This magnitude and the total amount of currency payments are necessarily the same, and the "equation of exchange," as it is called, is the algebraic expression of this relation.

It is P which measures the general level of prices and (inversely) the value of money, and the effect upon P of any of the other magnitudes is clearly shown by the equation. Changes in M or M' , or V or V' tend to cause similar changes in P ; changes in T tend to cause opposite changes in P .

Significance of the quantity theory. Few principles of economic science have been responsible for so much learned dispute and so much popular misunderstanding as the theory of just what causes determine the general level of prices and, more specifically, what the relation is between the quantity of money and the price level. The battle has raged most fiercely about the so-called quantity theory of money and the equation of exchange. While anything like exhaustive study of this problem lies necessarily outside the limits of this elementary book, we can with profit undertake some further analysis of the quantity theory and its algebraic expression, with the purpose of seeing clearly just what the principle means and some at least of the things that it does not mean.

The equation of exchanges sets forth the necessary balance, over a year or some other limited period of time, between six specified factors. The quantity theory of money expresses the influence upon the price level of each of the other factors as a "tendency"; *i.e.*, as the change in P which would follow a given change in any one of the other factors, "other things being equal." But other things

¹ Strictly speaking, $M'V'$ includes also any other forms of currency that may have been used and which, for simplicity's sake and because of their practical insignificance in the problem, we shall hereafter ignore.

may not — very generally will not — be equal. For example a decrease in the quantity of money (M) would decrease the magnitude on the left side of the equation; this would require an equal decrease on the right side, necessitating a decrease in the price level (P) — always assuming no change in any other factor. But a decrease in M might be counteracted by an increase in M' , necessitating no change in P , or even leading to an increase in P . Changes in M might likewise be counteracted by opposite changes in V or V' . Again a decrease in the quantity of money might fail to cause the price level to fall because of a simultaneous decline in the volume of trade (T), which would make the necessary adjustment on the other side of the equation. Many other combinations of changes are possible. Failure to understand this mutual balance of the factors involved leads to many erroneous assumptions, the most common and pernicious being the naïve notion that any change in the quantity of money necessarily causes a similar and even proportional change in the price level. While this is in truth the tendency, history records many instances where an increase in the quantity of money has not increased the price level or has not increased it proportionally, and *vice versa*.

Again, for the sake of clear thinking, it must be noted that the several factors whose influence upon the price level concerns us are not entirely independent of each other. Changes in these factors may, in part, work themselves out on each other rather than on P . We have already noted the relation which may exist between money and bank deposits (M and M'); so far as this relation is concerned the consequence is to strengthen the influence of changes in M or M' on P .

An opposite consequence follows from the fact that a sudden great increase in the quantity of money might have an effect on velocity of circulation, money passing on the average more slowly because of the greater quantity in people's possession. A sudden great decrease might have the opposite effect, causing the limited quantity of money to be turned over more rapidly. As regards the relation between M and M' and T , it is true that a sudden great increase in M or M' would very probably increase temporarily the purchases of the people and so increase T . A corresponding de-

crease in M or M' might decrease T . These relationships, tending to weaken the effect of changes in M and M' upon P , would generally be temporary in their effects, ultimately yielding to the habits of the people and other more fundamental influences.

Even more important is the relation between the volume of trade and the quantity of currency. There can be no question that an increase in the quantity of goods exchanged is very likely to have a material effect in increasing the amount of money and still more the amount of deposits, while a decrease in trade has the reverse effects. As we have seen, an elastic currency responds to business needs, especially through changes in the amount of bank notes (which are money) and bank deposits. Changes in T thus may cause similar changes in M and M' , and such changes will offset the tendency of the changes in T to cause opposite changes in P .

Finally we must note — although limitations of space prevent its demonstration — the fact that the equation of exchange is chiefly useful in portraying certain *long-run* relationships. It is less useful in explaining *short-run* changes. This is in part due to the fact that the equation combines all transactions in one total (T) and averages all prices so as to give one price level (P), whereas it is the various relationships between separate prices, rather than their average as shown by the general price level, which are of chief significance as causes of the short-run disturbances which are constantly appearing in the economic system.

As thus elaborated and qualified, the quantity theory of money is generally accepted as a true statement of the long-run relationships between the quantity of currency and the general price level.

Practical limitations of the quantity theory. While its theoretical validity is thus established, the practical utility of the quantity theory of money, either as a means of predicting price changes or as a guide to official control of the price level, is by no means so great as is sometimes assumed. This is due both to the interdependence of the several factors in the equation of exchange, to which our attention has just been directed, and also to the practical difficulty of discovering the exact magnitudes of the various factors. With the exception of the quantity of money and the general price level,

these magnitudes must generally be supplied by means of estimates and calculations which are subject to a wide margin of error.

Historical evidence from American experience. Both the theoretical complexities and the practical limitations of the quantity theory place obvious difficulties in the way of its historical verification. Nevertheless the simpler and more fundamental relationships between the quantity of currency, the volume of trade, and the general price level may readily be verified in a broad way by reference to historical changes.

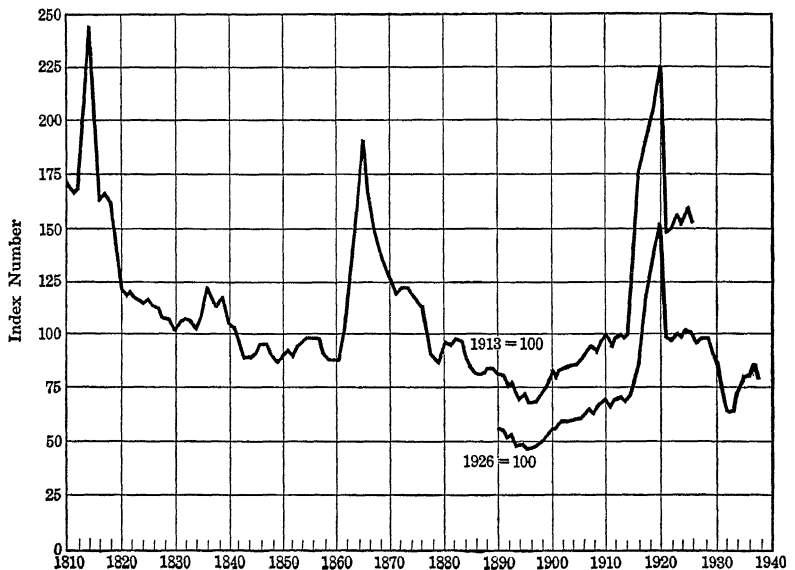


FIG. 38

Many of the incidents referred to in previous chapters on money will now be recognized as illustrations of this law. It will suffice for our present purpose to cite only a few examples, taken from the monetary history of the United States since 1860. As an aid to this inquiry we have combined several series of index numbers to give the graph in Figure 38. It will be seen — to go no further back than the Civil War — that prices rose sharply from 1861 to 1864, fell from 1864 to 1897, rose from 1897 to 1920, especially during the World War, fell precipitously in 1920 and 1921, remained

comparatively stable from 1921 to 1929, took a long downward slide in the first three years of the depression, and recovered to some extent from 1933. The most pertinent available monetary and trade data accompanying this succession of rises and dips in the price level are summarized in the following sections.

An increase in the quantity of money. Irredeemable money (1861-1864). At the end of the fiscal year 1861 (that is, June 30, 1861) the total monetary stock of the United States, as calculated by the Treasury statisticians, was 488 million dollars.¹ During the next three years there were issued, in connection with the financing of the Civil War, 447 million dollars of United States notes or greenbacks, 23 millions of similar notes of denominations less than one dollar, called "fractional currency," and 169 millions of various other forms of government fiduciary money. These issues, in combination with certain other monetary changes, brought the total monetary stock to 1,063 million dollars in 1864. Thus in three years the quantity of money was more than doubled. Bank deposits at this period were not an important part of the currency system, and in any event we have no information as to their amount sufficiently reliable for our present analysis.² In view of the great in-

¹ In the present discussion the figures of total monetary stock of the United States do not include the "minor coins" (figures for which are not readily available for the earlier years), though they do include the token coins of silver. For a more refined investigation it might be desirable to use, not the total monetary stock, but only such money as is in "circulation" or is the basis of other currency in circulation. However for the sake of the broad conclusions which we are seeking, the use of figures of the total monetary stock is sufficiently accurate.

² The figures for the monetary stock used in this historical sketch are from the published reports of the United States Treasury Department and are satisfactory for our purposes. Similar reliable data as to individual bank deposits subject to check are not available. This is due chiefly to uncertainty as to the distinction between time and demand deposits in the statements of the banks and to frequent changes in the official definitions of these terms. These circumstances make the figures published by the United States Comptroller of the Currency unsuitable for our present use. Much careful study has been given to this matter by certain economists, who have made their own estimates of the amount of deposits subject to check in selected years. Although these estimates, on account of the nature of the material, must be regarded as only moderately reliable, they have been used at those points in the present discussion where it has seemed worth while to include deposits in the analysis. Fisher's estimates have been used for the years 1896-1915; those of Snyder, for the years 1919-1920; and Currie's estimates for the period 1921-1935. Cf. Irving Fisher, *The Purchasing Power of Money*, 1911, and *American Economic Review*, June, 1916, pp. 457-458; Wesley C. Mitchell, *Business Cycles*, 1928, p. 126 (for Carl Snyder's figures); Lauchlin Currie, *The Supply and Control of Money in the United States*, second edition, 1935. The estimates for 1934 and 1935 were furnished to the authors by Mr. Currie.

crease in money, the price movement is what we should expect. The wholesale price level nearly doubled, rising from 88 in 1861 to 167 in 1864. Reciprocally the purchasing power of the dollar declined from 114 to 60.

The greenbacks were from the first irredeemable paper money. Except on the Pacific coast, they expelled most of the gold and silver from circulation and were themselves the basis of the monetary system. They depreciated in terms of gold, the depreciation being generally fairly proportional to the rise in the general price level. For example it required during 1864 on the average 203 dollars in greenbacks to buy 100 dollars in gold coin or an equivalent weight of gold bullion.

A decrease in the quantity of money (1864-1868). During the next four years conditions were reversed. There were reductions in the volume of greenbacks, state bank notes, other forms of fiduciary money, and gold coin. These were offset in part by increases in national bank notes and fractional currency, the net result being a decrease in the total monetary stock from 1,063 million dollars to 888 millions. The decline in the quantity of money tended to lower the general price level, an effect which was also promoted by the growth in trade after the war. The joint effect was a drop in the index number from 167 in 1864 to 142 in 1868. Considered as a change in the value of the dollar, this indicates an advance in its purchasing power from 60 to 70. The depreciation of the paper money in terms of gold also became less. In 1868 it took 140 dollars in greenbacks to buy 100 dollars in gold, whereas it required 203 dollars in 1864.

An increase of trade (1868-1879). The effect of increased trade stands out even more clearly in the history of the next decade. From 1868 to 1879 the total quantity of money remained fairly stable as compared with the preceding years. There was a gradual increase, from 888 to 1,034 millions. This increase in the quantity of money, other things being equal, would presumably have caused a moderate rise in prices. But other things were not equal. The country was growing rapidly in population and wealth, and trade was expanding; the increase during this period was so great as to outbalance the opposite tendency of the increase of money. Hence

the net result was a fall of prices. From 1868 to 1879 the index number of wholesale prices fell from 142 to 85. Reciprocally the purchasing power of the dollar rose from 70 to 118.

During this period the greenbacks continued their rise toward par in terms of gold, which was finally reached in the closing days of 1878, and on January 1, 1879, the government began redeeming them in gold on demand. Thus ended the irredeemable money régime which had continued for seventeen years; the metallic standard was restored. By this time the volume of greenbacks had been permanently fixed at 347 millions, and the fractional currency, state bank notes, and miscellaneous forms of fiduciary currency had been eliminated from the monetary system.

Increases in money and trade (1879-1897). The next chapter in our history takes us to about the year 1897. During these eighteen years the monetary stock nearly doubled, being 1,907 million dollars at the close of 1897. The increase was due to the silver dollars of the Bland-Allison and Sherman acts and to a material gain in gold, overcoming a decline of nearly 100 millions in national bank notes. It was in this period that bank deposits became an important part of the currency system, overtaking money in quantity; it is estimated that by 1897 the total of checking deposits somewhat exceeded two and a half billion dollars. This gain in the quantity of money and deposits was, however, after the first few years, not enough to counteract the phenomenal expansion of the country's trade which, gathering momentum after the Civil War, progressed beyond this period. These two influences working against each other in the equation of exchange, the net result was a quick rise in prices in 1880 and thereafter a steady decline, which was not permanently checked till after 1897. The index number of prices in this year was 67, which marks the low record to the present day. The purchasing power of the dollar in 1897, on the basis of 1913, was 149.

Monetary inflation through gold. The period from 1897 to the present has been on the whole one of monetary inflation. Up to the beginning of the World War in 1914 the inflation, strangely enough, was not due to paper money issues but to gold, in marked contrast with previous and subsequent periods of monetary expan-

sion. True the circulation of the national banks increased from 231 millions in 1897 to 759 millions in 1913, but this was a moderate increase alongside the extraordinary growth of the gold currency from 696 to 1,871 millions. This was the result of a tremendous gain in the output of the world's gold mines, a phenomenon of the utmost importance in view of the well-nigh universal adoption of the gold standard. We can well afford to give some attention to the history of the production of the world's chief monetary metal.

The world's production of gold. It is estimated that the average annual production of gold throughout the world four centuries ago (from 1493 to 1520) was a little less than 4 million dollars' worth, and that during the next three and a quarter centuries the rate of production increased gradually to about 13 millions in the decade 1831-1840. Then came an unexampled increase, due principally to the rich discoveries in California and other western states, which in twenty years had multiplied the world's annual production tenfold; it averaged 134 millions in the five years, 1856-1860. After this climax production fell off gradually to a little less than 100 millions in 1881-1885. Then began a new development, resulting from fresh discoveries of gold in South Africa, Canada, and Alaska, and to improved processes of extracting the gold from low grade ores. Thus in a few years the world's annual production reached a new high level, which has been substantially maintained ever since. In the first year of the twentieth century, 261 million dollars' worth of gold were added to the world's stock. In 1906 the year's production exceeded 400 millions, and in only three years since (1921-1923) has it been substantially below this figure. In 1929 there began another spurt, which carried the year's production above 500 millions in 1933. For the past twenty years the world's store of gold has been accumulating at more than thirty times the rate at which it was being built up during the eighteenth and first half of the nineteenth century. The gold produced in each of the years from 1908 to 1916 and 1930 to 1936 exceeded the total production of the first forty years of the nineteenth century.¹

¹ These facts are shown in more detail in the table on pages 667-8.

Effect on the price level (1897-1913). This deluge of gold into the world reservoir had its effect on the monetary system of the United States. In the sixteen years from 1897 to 1913, the total stock of money doubled again (rising from 1,907 millions to 3,720 millions), the three principal items in the increase being: subsidiary silver coins, 99 millions; national bank notes, 528 millions; and gold, 1,175 millions. During this period also the use of bank deposits was increasing very rapidly; it is estimated that the total of individual deposits subject to check rose to about eight billion dollars by 1913. These additions to the currency stock were more than enough to counteract the continuing growth in the country's trade, and the result was a turn in the movement of prices. The general price level, which at the end of a long recession had reached its low point in 1897 (index number, 67), now began a rise which carried it, almost without interruption, to 100 in 1913 (the base year of the series). This meant a fall in the purchasing power of the dollar from 149 in 1897 to 100 in 1913.

Monetary inflation of the World War (1913-1920). The World War ushered in a period of inflation exceeding anything in the previous history of the United States. The steady outpouring of the world's gold mines continued, but now the United States began to

WORLD'S GOLD PRODUCTION

Annual average for the period

PERIOD	FINE OUNCES	VALUE
1831-1840	652,291	\$ 13,484,000
1841-1850	1,760,502	36,393,000
1851-1855	6,410,324	132,513,000
1856-1860	6,486,262	134,083,000
1861-1865	5,949,582	122,989,000
1866-1870	6,270,086	129,614,000
1871-1875	5,591,014	115,577,000
1876-1880	5,543,110	114,586,000
1881-1885	4,794,755	99,116,000
1886-1890	5,461,282	112,895,000
1891-1895	7,882,565	162,947,000
1896-1900	12,446,939	257,301,100
1901-1905	15,606,730	322,619,800
1906-1910	20,971,575	433,520,960
1911-1915	22,213,810	459,200,175
1916-1920	18,966,758	392,074,674
1921	15,974,962	330,231,792
1922	15,451,945	319,420,063

WORLD'S GOLD PRODUCTION (*Continued*)

PERIOD	<i>Annual average for the period</i>	
	FINE OUNCES	VALUE
1923	17,790,597	\$ 367,764,279
1924	19,031,001	393,405,653
1925	19,025,942	393,301,128
1926	19,349,118	399,981,749
1927	19,397,757	400,987,213
1928	19,399,124	400,995,484
1929	19,585,536	404,968,955
1930	20,836,318	430,724,934
1931	22,329,525	461,592,277
1932	24,150,761	499,240,663
1933	24,962,408	516,018,675
1934	27,372,374	958,033,090*
1935	29,999,245	1,049,973,580
1936	32,960,158	1,153,605,530†

* Value at \$35 per ounce in 1934 and thereafter.

† Figures for 1936 subject to revision.

receive more than her normal share. For eleven years (1913-1924) the gold coin and bullion in the United States monetary system grew at the average rate of 238 million dollars a year. Whereas at the end of 1913 the United States held a little less than two billion dollars of gold out of a total world stock of about eight and a quarter billions (even then far exceeding any other nation), it has been estimated that at the end of June, 1924, the United States held the huge total of 4,488 millions, having more than doubled her holdings and having gained possession of nearly half the total gold stock of the world.¹ The American gold stock alone in 1924 was considerably more than the total of all kinds of money ten years

¹ The following table, based principally upon the reports of the Director of the United States Mint, shows the result of the World War upon the gold money of some of the important nations:

<i>Gold money held by</i>	<i>December 31, 1913</i>	<i>About Dec. 31, 1923</i>
Russia	\$1,011,500,000	\$ 45,043,000
Germany	915,700,000	119,300,000
Canada	142,500,000	227,292,000
Italy	265,000,000	215,697,000
Holland	60,900,000	233,876,000
Argentina	292,600,000	472,161,000
Spain	92,500,000	487,687,000
Japan	64,963,000	602,188,000
France	1,200,000,000	709,479,000
Great Britain	830,100,000	759,174,000
U. S. A.	1,904,700,000	4,247,201,000
World total	\$8,240,000,000	\$8,925,000,000

earlier. Upon this foundation of gold was erected an unprecedented superstructure of credit money and deposit currency. Federal reserve notes, which first appeared in the monetary system in 1914, reached a total of 3,406 million dollars in 1920. The total monetary stock rose from 3,720 millions in 1913 to 8,066 millions in 1920. Deposits subject to check increased during these seven years from eight billion dollars to about twenty billions.¹

If the reader will turn back to the series of annual index numbers on page 664 he will find the record of soaring prices which we should expect as a result of such an increase in the quantity of currency. By 1918 the price level had doubled; it reached a maximum of 226 for the year 1920. The highest monthly index number recorded was 247 for May, 1920.

Lower prices since 1920. The crisis of 1920 brought a sudden collapse of prices, the index number dropping to 147 for the year 1921. We must reserve for a later chapter our study of the manner in which business crises operate to liquidate checking accounts along with commercial loans and to decrease the velocities of circulation of currency; at that place we shall observe also how the sharp rise in unemployment reduces purchasing power and so intensifies the fall in prices, in spite of a decline in volume of trade. It is such influences as these that chiefly account for the phenomenal drop in the price level in 1920 and 1921.

During the period between the crises of 1920 and 1929 there was only a moderate increase in the country's monetary stock, which reached 8,418 million dollars on June 30, 1929. This increase of only a little over a third of a billion was all in gold; in fact there was a decrease of over a billion dollars in other forms of money, sufficient to counteract an increase of nearly a billion and a half in gold. This influx of gold went almost entirely into the possession of the federal reserve banks and the United States Treasury, where, in accordance with a conservative policy, it was to a considerable extent held as an idle store and not permitted to have the effect upon the money in circulation which might otherwise have occurred. Changes in the monetary stock during this decade must there-

¹ For this figure, we have taken Snyder's estimate and reduced it by one billion to conform with the estimates of Currie, which are used from this point.

fore have been approximately neutral in their effect upon the general price level. Bank deposits subject to check, after dropping sharply in connection with the crisis of 1920, rose from about 18 billions on June 30, 1921, to $22\frac{3}{4}$ billions on June 30, 1929.

This period — from 1921 to the panic of 1929 — was one of relatively stable prices, the index numbers remaining close to 150 throughout. This new level was just about fifty per cent higher than prevailed during the five or six years immediately preceding the World War. The purchasing power of the dollar in 1920 was only 44 (as compared with 100 in 1913 and 102 in 1914) and from 1921 to 1929 it stood at about 67.

In the fall of 1929 came another crisis, succeeded by a period of economic depression, during which the price level declined to a low point of 98 for the year 1932. During these three years the monetary stock was increased by about half a billion dollars — to 8,878 millions on June 30, 1932. Whatever inflationary tendency this increase may have had was more than counteracted by the ordinary conditions of panic and depression. In particular bank deposits subject to check declined from $22\frac{3}{4}$ billions on June 30, 1929, to less than sixteen billions on June 30, 1932, and the velocity of circulation of bank deposits, which was estimated at about 45 for the years 1925–1929, suffered a drastic slowing down to 26 by the end of 1930 and reached a low of 16 in the last quarter of 1932.

Irredeemable money and prices (1933–1938). Beginning in March, 1933, the American monetary system underwent revolutionary changes, with the principal features of which the reader is now familiar. An irredeemable money system displaced the gold standard, and the nominal monetary unit (the gold dollar) was reduced 41 per cent in weight. These modifications complicate the task of following the changes in the total monetary stock. As officially reported, the stock (still excluding the minor coins) had grown to 20,867 million dollars by September 30, 1938. This nominal increase of nearly 12 billions, or 135 per cent, since June 30, 1932, included an increase of 1,497 million dollars in federal reserve notes, offset by a decrease of 515 millions of national bank notes; silver was responsible for an increase of about a billion. The largest item is the increase in gold — nearly 10 billions — the

combined result of gold purchases and devaluation of the dollar (making every 100 dollars of gold into 168 dollars). So far as the equation of exchange is concerned, at least that part of the Treasury's gold in excess of the amount of gold certificates should be excluded from the monetary stock. Such exclusion would make the total monetary stock 18,102 million dollars as of September 30, 1938, which is probably the amount most nearly comparable with the total monetary stock as of earlier dates. Bank deposits subject to check have increased gradually since their decline to about 16 billions in 1932. They were estimated at 25 billions at the close of the year 1937, an increase of 9 billions in these five years.

During this period — since March 4, 1933 — it was the avowed purpose of the Administration to raise prices, the general level of 1926 being accepted as the goal. The various changes made in the monetary system evidently had this purpose. The same purpose was back of other measures taken, such as open-market purchases of government securities by the federal reserve banks for the sake of expanding bank credit, relaxing the reserve requirements of the federal reserve banks, and official pressure upon the banks generally to lend freely at low interest rates and with less exacting standards as to security.

The policy of the government can hardly be said to have achieved success. The price level rose only moderately, reaching 78.1 (on the 1926 base) in August, 1938, as compared with 64.8 in 1932 and with the desired goal of 100 as in 1926. The last word on the effectiveness of the government's monetary policy in recent years has not been said. Throughout this chapter attention has been called to the fact that the relation between the quantity of money and the price level is a long-run relationship. There can be no doubt of the inflationary character of most features of the government's monetary and banking policy. The fact that the desired price level was not quickly attained must not be regarded as proof that the measures taken may not have released influences which in the course of time may lead to great inflation and a soaring price level. Further analysis of the very complicated forces which have acted upon the price level in recent years will come with the investigation of the problem of trade cycles in a later part of this book.

Far-reaching effects of monetary fluctuations. Fluctuations in the value of money are fraught with the most serious consequences to all the people, inflicting fortuitous loss upon some and bestowing undeserved gain upon others. Rights to future payments are normally expressed in terms of the country's monetary unit. When the time for payment comes the debtor will pay and the creditor will receive money whose value may be quite different from that which existed when the agreement or contract was made. Suppose that a farmer borrows \$10,000 at a time when wheat is worth \$1.00 a bushel. If before the time for payment there has been such a fall in the value of money that prices, including the price of wheat, have generally doubled, the farmer will be able to pay his debt in dollars worth only half as much as those he borrowed. When he made the loan, it required the sale of 10,000 bushels of wheat to produce \$10,000. Now he need sell only 5,000 bushels to obtain the \$10,000 needed to pay his debt. The debt has been virtually halved, to the advantage of the farmer and of course to the corresponding loss of the one from whom he borrowed. On the other hand, if in consequence of a doubling of the value of money the price of wheat had fallen to fifty cents the result upon both parties would have been the reverse.

All investments which consist of rights to receive stated sums of money, such as bonds, mortgages, annuities, insurance, etc., are subject to the same principle. The investor is compelled in spite of himself to gamble on the value of the dollar. If it goes up, the dollars he receives will buy more; if it falls, his dollars will buy less. Such investors were fortunate gainers from this cause during the long period of falling prices that ended in 1897. After that the most conservative investors saw their resources dwindle from year to year as the index number of prices rose. The man who in 1904 put money in the savings bank, where he left it to accumulate at five per cent till 1918, suffered a loss through the shrinkage of the dollar sufficient to wipe out all the gain from his compound interest for the fourteen years. On the other hand, a period of drastic price decline may injure investors through the failures and bankruptcies of many business concerns whose securities they hold.

Those whose incomes are in the form of salaries and wages are affected by fluctuations in the value of money in the same way as are creditors. They have agreed to render services in return for a stated number of dollars, which as they are received may be worth more or less than when the agreements were made. Only as agreements expire and new ones are made can the parties undertake to make new terms conforming to the new monetary value, and these new agreements are equally subject to the vicissitudes of future changes.

It is thus that adjustment lags behind the fluctuations of the monetary standard on all agreements except those which are for immediate payment or for relatively short terms. Here there is the current adjustment, through raising and lowering prices, which produces those fluctuations in the general price level by which the community adjusts itself as best it may to the changing monetary standard.

In inflationary periods (periods of rising prices and falling value of money) debtors and all those who have contracted to pay fixed sums of money gain at the expense of all creditors. Those who gain in such periods include railroads, utilities, and industries with heavy bonded indebtedness, farmers and home owners whose properties are mortgaged, and speculative borrowers who borrow money in order to buy goods for the express purpose of selling them later and thus profiting from the inflation. Business men generally reap an advantage, since the prices they obtain for the goods they sell tend to rise more rapidly than the costs that they have to meet. On the other hand, all bondholders and recipients of fixed rents suffer loss, together with all owners of savings accounts and insurance policies and all those with rights to annuities payable in money. All workers for wages and salaries lose in so far as their pay is slower to rise than the cost of the things they buy, but it must not be lost sight of that a period of rising prices is usually a period of full employment and usually the income of labor as a whole will be increasing because of expanding employment and rising wage rates. In so far as this is true, the losses of those individuals previously employed whose incomes will purchase less than before tends to be offset by the gain of those previously un-

employed who find employment. To some extent, the rising tide of prosperity which usually accompanies rising prices will offset the losses even of bondholders and owners of savings accounts. For such a period is likely to see the assets of insurance companies and banks rise in value and will always see many previously shaky companies restored to financial stability. For all classes in the community a period of rising prices is likely to be one in which a higher cost of living is compensated by greater personal security.

All this is reversed in deflationary periods (when money rises in value and prices fall). Now lenders as a class gain, and "the bloated bondholder" replaces "the profiteer" as the target for popular attack. Workers and their families gain at first in that their pay checks are slow to shrink, but because the fall in prices and the growing debt structure adversely affect business profits, unemployment increases, balancing against the gains of those who are still employed the losses of those who cannot find work. Even the "bloated bondholder" may eventually find himself holding worthless "securities."

Monetary reform. It will generally be taken for granted that stability of value is a desirable feature in any monetary system. Stability of value of money means a general price level correspondingly stable. Some would go so far as to hold up as the ideal a monetary unit whose purchasing power would be always the same. In the past those who have understood the economic principles of money and have sought the maintenance of the most stable standard possible have generally had their hands fully occupied in repelling the attacks of those who sought to displace the gold standard by something even less stable. More recently the widespread abandonment of the gold standard and the present uncertain state of the monetary systems of the world have directed much thought to the fundamentals of monetary theory and the problems of monetary reform. Is a stable monetary standard possible? How may such a standard be obtained? The limits of this elementary book will not permit us to explore this problem much beyond the mere outlining of some of the more significant current suggestions for a stable monetary system.

There are first those plans which would adhere to the gold standard and look to such regulation, official or otherwise, of the amount of credit money — generally bank notes — and the amount of bank deposits subject to check as would regulate the value of currency through control of its quantity. Such plans have undoubted theoretical merits, while at the same time involving many practical difficulties and facing serious political hazards.

Some years ago Irving Fisher proposed his now well known "commodity dollar" or "stabilized dollar." Adhering to the gold standard, this plan would make the monetary unit — the dollar — not a certain weight of gold as now, but a variable weight of gold, so adjusted from time to time as to keep its purchasing power virtually constant. Gold coin would not be used in actual circulation. The government would be as now the custodian of the country's monetary gold. There would be in circulation gold certificates and other forms of money reading in terms of dollars (not necessarily much different from the present forms), all of which would be redeemed in gold by the government on demand. The amount of gold to be given for a dollar would be officially determined from time to time on the basis of the index number of prices, so that a gold dollar would always purchase about the same quantity of goods in general. Variations of individual prices with respect to each other, like the movements of the separate bees in a swarm, would in no wise be affected. But the movements of all prices in general would be stopped; the swarm as a whole would become stationary. This plan, especially if simultaneous adoption by most of the leading nations be assumed, might conceivably operate successfully under normal circumstances. Whether political obstacles would prove insurmountable is a serious question.

There is reason to doubt whether either of the foregoing plans would be strong enough to withstand the forces tending to a drastic decline of prices in times of major crises. On the other hand, the operation of such a plan as the "stabilized dollar," though not perhaps able to cope with violent short-run price changes, might have considerable influence in smoothing the long-run disturbances which are known as "long waves." Elimination or moderation of these long-range fluctuations might have a salutary effect on the

shorter fluctuations of the business cycle. In that case, the strain of the crisis itself would be easier to withstand.

Managed money. During recent years there has arisen a school of thought which holds that the ideal monetary system should not rest upon standard money at all, but should be a regulated irredeemable paper money system. The value of the monetary unit would be controlled by official action, for which the two most effective instruments are (1) control of the quantity of money and of bank deposits so as to maintain a uniform or otherwise controlled level of domestic prices, and (2) dealings in foreign exchange in order to maintain the value of the money in relation to the monetary units of other countries.

There has been of late some experience of this sort of monetary control, especially with reference to foreign exchange rates. Since their recent departure from the gold standard the leading nations have instituted measures to stabilize the values of their paper currencies in the international exchange market. For example the British government undertakes to keep the value of the pound in terms of dollars, francs, etc., fairly stable. This is accomplished by purchases and sales of gold in the open market. These operations require a fund set aside for this particular purpose. Great Britain in 1932 established the Exchange Equalisation Account, as its stabilization fund is technically known, for the purpose of stabilizing exchange rates and "managing the currency." The fund is controlled by a committee of the Treasury and the Bank of England. It is an immense fund, approximating 575 million pounds and consisting of gold, foreign exchange, and national currency, and is secretly managed. To take a concrete case, let us suppose that in England the value of the pound begins to decline in terms of the dollar. The British committee may halt the decline if it is able to buy pounds and sell dollars; in other words, if it offers to redeem pounds with dollars at some price. But if it is to redeem pounds in terms of dollars, the equalization account must either have accounts in the United States with private banks or with the federal reserve banks from which it can draw dollars, or else it must possess sufficient gold so that it may secure dollars from the United States Treasury by selling gold to the Treasury.

Similar machinery and stabilization funds have been set up in the United States, France, Belgium, Holland, and Switzerland. The United States, as the reader will recall, devalued the dollar at the time of departure from the gold standard, thereby acquiring a "profit" of nearly 3 billion dollars, from which her stabilization fund of 2 billions was taken. Great Britain, on the other hand, raised her 575 million pounds by an issue of securities (*i.e.*, by a loan). These stabilization funds have become a fundamental feature of the currency systems of these countries, and certain authorities are disposed to favor such control as a permanent part of the financial machinery.

In recognition of such control as this, it has become the habit of late to apply the term "managed money" to the recently adopted irredeemable paper money systems. To avoid misleading inferences it should be remembered that not all irredeemable paper money systems have been characterized by any important degree of "management," while, on the other hand, there may be a good deal of "management" in connection with a gold standard system, through control of discount rates, "open-market operations," etc. There was for example far less management of the Continental bills of credit or the greenbacks than there was of the United States gold standard system in the twenty years before 1933. It is a fact however that the leading nations which went off the gold standard in connection with the crisis of 1929 and the following depression have exercised a considerable control over the international values of their paper money currencies.

In support of "managed money" there has been marshalled a considerable weight of theoretical argument based upon the quantity theory of money. There is theoretical argument also against the plan. In particular, the hazards of a monetary standard controlled by government and so always a potential football for political contest are not to be overlooked. The temptation to pay the costs of government by paper money issues and bank credit, in lieu of the less popular taxation, and the popular sentiment in favor of cheap money which is so prevalent in many countries furnish two powerful forces which as a matter of fact have eventually directed most irredeemable money systems down the path of uncontrolled

inflation. It may be observed that these dangers are especially to be feared in the United States. Historically the world's experience with irredeemable money has been, almost without exception, disastrous. On the other hand, Great Britain, the United States, and a large number of other nations have now maintained irredeemable paper money systems, involving various degrees of "management," for about a decade at least without bringing disaster.

Before leaving this topic it should be noted that monetary theory is not wholly clear even as to what should be the aim of a managed money system. The idea of a constant price level is not so simple as may at first appear. There is reason to believe that in our existing economic order a stable price level might conceal the presence of factors making for eventual disequilibrium. Neither can it be said that the value of the monetary unit ought to be maintained at any specified level in terms of gold. Since Great Britain went off the gold standard in 1931, the gold value of the pound has in fact fluctuated. It is proposed by some that the managed money system be employed, not to preserve a stable price level, but to regulate the price level so as to bring about desired ends embraced in the government's program. This takes the problem into a realm in which economic science does not yet speak with sureness. For the further study of these important problems the reader is recommended to more advanced and special treatises. Our investigation thus far would seem to justify the conclusion that, in spite of the recognized defects of the gold standard and the attractive features of some of the current monetary proposals, as a matter of practical policy the presumption is still strongly in favor of the traditional gold standard.

EXERCISES

1. Compute an index number of prices for each of the three years, on the 1926 base, using the simple average of price relatives:

	1926	1932	1938
Wheat (bu.)	1.20	.40	.60
Cotton (lb.)15	.06	.09
Coal (ton)	5.00	4.00	7.00
Iron (ton)	12.00	9.00	15.00
Petroleum (bbl.)	1.20	.60	1.00

2. What is the change in the value of money if prices rise 50 per cent? If prices rise 25 per cent? If prices fall 20 per cent? If prices fall 50 per cent?
3. In terms of the "equation of exchange" explain what effect each of the following would tend to have upon the general level of prices in the United States:
 - (a) The issue of a large amount of paper currency by the federal government.
 - (b) Increased productive efficiency of American industry.
 - (c) Open-market purchases of government securities by the federal reserve banks, increasing the size of commercial bank reserves.
 - (d) A decrease in borrowing by business men.
 - (e) Improved facilities for transfer and collection of checks.
 - (f) Expectation of rising prices.
 - (g) A law increasing the size of the reserve ratios which banks must maintain.
 - (h) Floods, droughts, strikes.
 - (i) Increased volume of loans by banks to business men.
 - (j) A sudden wave of thrift, or an increased desire to save.
4. (a) In the last stages of inflation in Germany after the World War, prices rose more rapidly than the increase in the quantity of money. Account for this situation.
- (b) In the three year period, 1930-1932, the quantity of hand-to-hand currency in circulation increased by more than 800 million dollars, yet prices declined severely. How, if at all, can this result be reconciled with the quantity theory of money?

SUGGESTIONS FOR FURTHER READING ON PART IV, MONEY AND BANKING

I. MONEY AND BANKING

- BRADFORD, F. A. *Money and Banking*. Second edition. New York, 1935
- WHITE, H. *Money and Banking*. New edition. Boston, 1935
- DOWRIE, G. W. *American Monetary and Banking Policies*. New York, 1930
- MOULTON, H. G. *The Financial Organization of Society*. Third edition. Chicago, 1930
- RUFENER, L. A. *Money and Banking in the United States*. Boston, 1934
- CURTIS, M., and TOWNSEND, H. *Modern Money*. New York, 1938
- ROLL, E. *About Money*. Toronto, 1934
- ROBERTSON, D. H. *Money*. New York, 1929
- BERNSTEIN, E. M. *Money and the Economic System*. Chapel Hill, 1935
- KEYNES, J. M. *The General Theory of Employment, Interest, and Money*. New York, 1936
- McGEE, J. D. *An Introduction to Money and Credit*. Revised edition. New York, 1933
- JAMES, F. C. *The Economics of Money, Credit, and Banking*. Second edition. New York, 1930
- GARIS, R. L. *Principles of Money, Credit, and Banking*. New York, 1934

- KILBOURNE, R. D. *Principles of Money and Banking*. Third edition. New York, 1927 (supplement by G. W. Woodworth, 1934)
- STEINER, W. H. *Money and Banking*. New York, 1933
- KEMMERER, E. W. *On Money*. Philadelphia and Chicago, 1934
- National Industrial Conference Board. *The New Monetary System of the United States*. New York, 1934
- CURRIE, L. *The Supply and Control of Money in the United States*. Second edition. Cambridge, Mass., 1935
- GREGORY, T. E. *The Gold Standard and Its Future*. London, 1932
- GAYER, A. D. *Monetary Policy and Economic Stabilization: A Study of the Gold Standard*. New York, 1935
- WESTERFIELD, R. B. *Our Silver Debacle*. New York, 1936
- MADDEN, J. T., and NADLER, M. *The International Money Market*. New York, 1935
- EINZIG, P. *The Sterling-Dollar-Franc Tangle*. New York, 1933
- Royal Institute of International Affairs. *The Future of Monetary Policy*. New York, 1935
- WESTERFIELD, R. B. *Banking Principles and Practice*. New York, 1924
- *Money, Credit and Banking*, New York, 1938
- PHILLIPS, C. A. *Banking and the Business Cycle*. New York, 1937
- HELDERMAN, L. C. *National and State Banks — A Study of Their Origins*. Boston and New York, 1931
- REED, H. *Federal Reserve Policy, 1921-30*. New York, 1930
- HARRIS, S. E. *Twenty Years of Federal Reserve Policy*. Cambridge, Mass., 1934
- CHAPMAN, J. M. *Fiscal Functions of the Federal Reserve Banks*. New York, 1923
- TIPPETTS, C. S. *State Banks and the Federal Reserve System*. New York, 1929
- ROBB, T. B. *The Guaranty of Bank Deposits*. Boston, 1921
- ECKHARDT, H. M. P. *A Rational Banking System (Branch Banking)*. New York, 1911
- SOUTHWORTH, S. D. *Branch Banking in the United States*. New York, 1928
- BENNER, C. L. *The Federal Intermediate Credit System*. New York, 1926
- WILLIS, H. P., and BOGEN, J. I. *Investment Banking*. New York, 1929
- CONANT, C. A. *History of Modern Banks of Issue*. Fourth edition. New York, 1909
- WILLIS, H. P., and BECKHART, B. H. (editors). *Foreign Banking Systems*. New York, 1929
- DEWEY, D. R. *Financial History of the United States*. Twelfth edition. New York, 1934 (see Table of Contents)
- NOYES, A. D. *Forty Years of American Finance*. New York, 1909
- *The War Period of American Finance*. New York, 1926

II. MONEY AND PRICES

- FISHER, I. *The Purchasing Power of Money*. New York, 1911
- MARGET, A. W. *The Theory of Prices*. New York, 1933
- WICKSELL, K. *Interest and Prices*. London, 1936

-
- LAYTON, W. T. *An Introduction to the Study of Prices*. London, 1914
- MITCHELL, W. C. *Gold, Prices, and Wages under the Greenback Standard*. Berkeley, 1908
- FISHER, I. *The Making of Index Numbers*. Boston, 1922
- MITCHELL, W. C. *The Making and Using of Index Numbers*. Bulletin of the United States Bureau of Labor Statistics, No. 284. October, 1921
- EDIE, L. D. *Money, Bank Credit, and Prices*. New York, 1928
- FISHER, I. *Stabilizing the Dollar*. New York, 1920
- SPAHR, W. E. *The Monetary Theories of Warren and Pearson*. New York, 1934
- FISHER, I. *Stable Money*. New York, 1935

INDEX

- Acceptance, 546
- Advertising, 184, 189-93; purpose of, 189, 191; a barometer of business prospects, 193
- Agriculture, and demand, 199; economic rent in, 351-75; government aid to, 269, 610, 630-32; level of prices in, 133, 285-86
- Amana Society, 98
- Assets, defined, 85
- Auctions, 207
- Balance sheet, explained, 546-50
- Bank credit, 76, 541-42, 560, 575-76; elasticity of credit, 570-71; expansion of, before 1929, 602-604; importance of, 576-77
- Bank holiday of 1933, 527-28, 612-13
- Banking, American, 578-608; American system in 1929, 597-600; American, since 1929, 609-35; system in United States in March, 1938, 628-29; branch, 582-86, 599-600, 621-22, 636, 640-41, 647; British, 640-47; Canadian, 647-53; centralized, 624-26; chain, 599-600, 622; criticism of government, 634-35; decentralized, 600-602; an economic index, 201; emergency legislation and its perpetuation, 612-16; Federal Reserve, 584-96; foreign systems of, 636-53; French, 637-40; functions of, assumed by United States government, 609-10; 630-35; history of American, 578-84, 609-35; legislation since 1929, 608, 610-28; political control, 625; principles of, 540-77; profit in, 561-62; situation in March, 1938, statement of, 629; weaknesses of American system, 600-604
- Banking act of March, 1933, 612-13, 621, 622, 624, 626
- Banking act of 1935, 614, 620, 621, 623, 626, 627, 628
- Bank notes, 502-503, 566-69; British, 642-44; Canadian, 648-49; defined, 566; discount for, 547-48; elasticity of, 570; federal reserve, 568, 587-91; French, 639; national, 581-82; regulation of, 568-70, 589-90, 636; relation of, to deposits, 566-69
- Bank of Canada, 650-53
- Bank of England, 641-47
- Bank of France, 637-40
- Bankruptcy, 214
- Banks, American, local (state), 578-79, 581, 582, 586, 591, 595, 618, 620, 621, 629; and RFC, 609-10, 612, 614, 615-16; capital requirements for, 620-21; central, 566, 586, 636-37, 651; failures, 607-608, 609; home loan, 610, 632, 633, 634; investment, 460-61, 622-23; land, 542, 610; Morris plan, 389, 620; private, 574, 581; mutual savings, 617-19, 624; savings, 542, 574, 581, 623-24; short-term loans by, 394-95
- See also* Commercial banks; National banks; Federal reserve
- Barter, 21, 498
- Bimetallism, 505, 511, 512, 534; in Europe, 515-16; in United States, 512; end of, in United States, 513-14
- Bland-Allison act, 517-18
- Bonds, explained, 83-85; interest payment on, 90; government, 580, 581-82, 589-90; valuation of, 331
- Borrowing, 542-45; by government, 381-82, 390, 391; for consumption and time preference, 381, 388-90, 391; for production, 381, 386-87, 405
- Brassage, defined, 532 n.
- Broker, note, 571
- Brook Farm, 98
- Building, and consumer demand, 199
- Business, forms of organization of, 74-97; liberty in, 188-89
- Business management, 36-37
- See also* Entrepreneur

- Capital, accumulation of, 31-34, 36, 56, 119, 383-86, 406; consumers, 30; defined, 29-30, 88 n.; a factor in production, 27; importance of, 324-29; producers, 30; production of, and time preference, 377-80; productivity of, 387-88; relation to income, 324-26, 332-33; valuation of, 324-32
- Capitalism, growth of, 53
- Capitalist class, 35-36, 400-402; defined, 35; essential service of, 402; power of, 35-36
- Capital stock, common, 82, 83, 84, 85; dividends on, 82-83, 90-91; explained, 81-83; preferred, 82-85; value of, 88-90; watered, 92
- Check, 551-58, 567-68; collection between banks, 553-55; definition of, 551-53; out of town, 557; payment by, 552-55; payment within bank, 552-53
- Classes, social, 429-36, 486-87
- Clearing house, 555-57, 595-96
- Coal, 57-58, 67, 68
- Coinage, 506; free, 506 n., 506-507; gratuitous, 507; ratio, 512 n., 512, 513; total, of silver dollars, 517-19; under bimetallism, 507-508, 512-14
- Collective bargaining, 419
- Commercial banks, 542, 574, 622-24; and affiliates, 602; collection between, 553-55; divorced from investment banking, 622-23; function of, 575-77; insured, 619; legislation concerning, since 1929, 609, 612-19; number and size of, 629; operations of, 542-71; profits in, 561; security holders of, 562-63; types of, 542
- Commercial paper houses, 571-73
- Commodity Credit Corporation, 615
- Competition, 117, 179-191; and advertising, 190; basis of modern industry, 112; and the buyer, 179 n., 183-84; classification of, 180; "cutthroat," 264, 316; defined, 179; importance of, 301-303; may degenerate into monopoly, 116; pure, and price, 180-81, 190-91, 229-31, 270-303; imperfect, and price, 183-87, 190, 233-36, 270, 305-22; non-price, 189; and producer, 187; requirements of, 179 n.; and sellers, 179 n.; unregulated, 70, 116; protects consumer, 302
- Consumer, balances marginal utilities, 153-57, 167; controls production, 113; and demand, 306; desires of, 194-95; incomes of, 195-98; influence of, on price, 136-37, 160-61, 204; protected by competition, 302
- Coöperation, 98-99; unconscious, 114-15; simple, 98-99; and division of labor, 99, 104-105
- Cooperative marketing associations, 630
- Corporation, 78-96; capital stock of, 81-83; forms of securities of, 81-85; income account of, 92-96; nature of, 79-81; property account of, 85; saving by, 384-86; significance of, 78-79
- Cost, analysis of, 214-16; average, 222-25, 241-43; bulk-line level of, 211; constant, 258-61; controls supply, 256-62; decreasing, 263-65, 293-96; defined, 15; differential, 222 n.; of production and discounted income, 335-37; effects of past, 213; and expanding output, 217-20; fixed, 214-19, 224-25; future, 212-13; increasing, 256-58 (*see also* Diminishing returns); individual, 220; limits prices, 210-13; long run and short run, 232-33; marginal, 238-41, 256-62; and output, 217-18; overhead, 215 n., 216; of production, 205 n., 205-10, 215, 229-47; relationship to price, 204-13; schedules and curves, 225-28; and the sellers, 204-28; unit, and large-scale enterprise, 219-20; variable, 214-19, 232, 243
- Cost accounting, 213
- Credit, commercial, 76; dependent on savings, 541; explained, 541-42; control of, 592, 625-28; function of, 575-77; government, 609-10; speculative use of, 624
See also Bank credit
- Credit money, 503
- Crisis of 1929, 606
- Crisis of 1931, 606
- Cultivation, margins of, 364
- Currency, elasticity of American, 500-501, 582, 588-89; elasticity of

- British, 644-46; elasticity of Canadian, 649; fractional, 673; need of elastic, 569-70, 582; quantity of, and price level, 663, 665-67, 668-81; quantity of, and value of money, 685; velocity of circulation and price level, 663, 667-68
- Customs barriers, 49, 56, 109
- Debt, and fluctuation of value of money, 682-84
- Deflation, 683
- Demand, and advertising, 191-93; analogous to supply, 252; ascertaining, 194; two concepts of, 276-79, 281; control of, 189-91; creation of, 202; controlled by marginal utility, 151, 153-60, 166-68; curve of, 162, 277, 306-307; defined, 161-62, 279; effect of changes in, 282-84, 293-95; effect of price and quantity, 165-66; elasticity of, 168-74, 243-44; factors of, 194; and imperfect competition, 243-45; importance of, 173; and income, 195; indication of future, 195-99; individual, 136, 153-61; law of, 164-68, 173-74; meanings of, 276-79; modification of, 202; not desire, 164; and price, 204, 245; and attainment of profit, 236; schedule of, 160-63; and seller, 176-202; 233-34; and sole buyer, 166-71; stimulation of, 189-95; influence of substitutes on, 307, 308-10; and supply, 204-205, 265; and supply, laws of, 273-75; total, 137, 160, 161-62, 163
- Deposits, bank, 544, 549-55; defined, 549; discount for, 550-51; elasticity of, 644-46, 668-71; individual, as currency, 539; insurance on, 617-20; interest on, 623-24; relation to bank note, 559-60, 566-68; velocity of circulation, 569, 668-69; volume of, 629, 667, 673, 680; British, 640-41, 644-46; Canadian, 649
- Depression, 132, 607-608; after 1929, banking legislation in, 609, 610-24; and interest rate, 400; and costs, 213, 232-33; and government policy of banking, 612
- Diminishing returns, in agriculture, 133, 353-58; importance of principle, 371; law of, 353, 354; and marginal cost, 355-57
See also Marginal cost
- Discount, 331-32, 544, 547-49, 550-51; bank, 542-43; explained, 544, 544 n.; by federal reserve banks, 586-87; and laws of price, 333-37; rate of, 564-66
- Distribution, 119; meaning of, 351; functional, 351, 467, 476-78, 480; personal, of income, 351, 470-75, 480-85; personal, of wealth, 473-76; unequal, 117-18
- Dividends, payment of, 82-83, 91-92; stock, 91-92
- Division of labor, 98-105, 418-19; economic effects of, 99-102; and the market, 102-103; a form of coöperation, 104-105; territorial, 103
- Dollar, commodity or stabilized, 684-86; devaluation of, 529-31; silver, after 1873, 517
- Draft, 546, 558
- Economic freedom, 119, 489
- Economic friction, 298-99
- Economic rent, 351-75, 454 n.; in agriculture, 351-55, 358-62; contractual, 352, 359-60; defined, 352; implicit, 352; and land value, 369; problems of, 371-75; relation to prices, 369-71; taxation of, 374-75; on urban lands, 365-66
- Economics, as a science, 8-9, 13; defined, 7-8
- Education, public, 70, 488-89
- Employment, during inflation, 683; indexes of, 197
- Endorsement, 545-46, 553, 554
- Entrepreneur, 36-38, 74, 112, 210, 458; and risk, 459, 460, 465
- Equilibrium, 270-73; law of, 275; price and market price, 275; and competitive prices, 313-17
- Equity, owner's, 87
- Exchange, defined, 21; development of, 49-50; equation of, 668-71; with money, 21-22, 49-50
See also Foreign exchange; Trade Exchanges, stock, 299-300
- Export-Import banks, 616
- Exports, of United States, 605

- Factory system of industry, 53-54, 60-63, 69-71
 Fair labor standards act of 1938, of United States, 448-50
 Fairs, 47-48
 Farm. *See* Agriculture
 Farm Credit Administration, 630-32
 Federal Deposit Insurance Corporation, 617, 618, 620, 623, 634
 Federal Farm Board, 269
 Federal Housing Administration, 634
 Federal National Mortgage Association, 634
 Federal Open-Market Committee, 626
 Federal reserve banks, 568, 584-97; deposits in, 591; discounting by, 586-87, 602; functions of, 586-87; gold holdings of, 587, 593, 594; and government business, 596-97; legislation concerning, since 1929, 609, 610-12, 613, 614, 616, 622-25; notes of, 503, 568, 587-91, 610-12; open-market operations, 591-92; rediscounting by, 586-87; reserves of, 593-94; statement of, 630-31
 Federal reserve system, 584-98; districts and member banks, 584-86; founded, 584; integration, 625; membership in, 586, 621; a national clearing house, 595-96; and national banks, 584; and open-market operations, 591-92, 626; organization of, 584-86; and panic, 594-95; powers of board, 622-27; and state banks, 620; statement of September, 1938, 630-31; and stock market, 609
 Financing, 540-42
 Fisher, Irving, and stable money, 685; on standards of measure, 661-62 n.
 Foreign exchange, rates of, 686-87; French dealings in, 638-39; standard, irredeemable, 539; under act of March, 1933, 613
 Freedom, personal, 50, 56, 106, 111; economic, 110-11, 118-19
 Free trade, at medieval fairs, 47-48; in England, 68, 109
 Gilds, 43-44, 45, 107
 Glass-Steagall act, 610-11, 612, 614
 Gold, 108; British dealings in, 644, 646, 647; bullion to settle international balances, 529-30, 639, 644; Canadian policy concerning, 652; clause concerning payment in, cancelled, 528, 531; embargoes on exports of, 527, 528; holdings of federal reserve, 587, 593, 594; free coinage of, 506-507; French holdings of, 640, 678 n.; hoarding, 536; international flow of, 108, 603, 644; nationalization of, 529-30; policy of purchasing, 529-30; production of, 676-79, 678 n.; ratios of, to silver, 507-15; as standard money, 505-507, 508; stock of, in United States, 527, 611, 676, 677-79, 678 n., 681; surrender of, in 1933, 528; Treasury's profit in, through devaluation, 530-31, 532, 590
 Gold bullion standard, 507, 519, 526, 625
 Gold certificates, United States, 503, 685
 Gold exchange standard, 505, 519-20, 605-606
 Gold reserve act of January 30, 1934, 529-32
 Gold standard, 505, 507, 518, 519, 527-29, 611, 684, 685, 687, 688; adoption of, by United States, 516, 518; desertion of, by United States, 525-26, 528-29; France and, 640; general desertion of, 606; Great Britain and, 644-46, 647
 Government, functions of, 110-11; American, attempts of, to raise price level, 681; attempts to control supply, 268-69, 285; American, banking functions assumed by, 609-10, 625, 630-35; borrowing of, 390-91; controls money, 499-500, 506
 Government notes, 502-503
 Government regulation, 111; in English history, 48-49, 106-108; in French history, 109; of industry, 489; of interest, 402-406; of price, 133-34, 489; of wages, 107, 448-51
 Greenbacks, 503, 513; régime, 517, 524-25
 Gresham's law, 509-10; operation of, 522

- Handicraft stage, 43-53
 Heaton, Hubert, on textile factories, 61-62
 Hoarding, 609-10, 612, 613, 619, 666
 Holding companies, 622
 Home Loan Banks, 610, 632, 633
 Home Owners Loan Corporation, 610, 633
 Hours of labor, 70-71
 Human beings, are they wealth?, 11-12; property in, 16, 18; services of, 15, 18-20, 26-27
 Income, of American families, 481-85; changes in, as a business barometer, 196-98; and capital, 332-33; controls production and consumption, 127-28; and demand, 195, 198-200; discounted, 335-37; functional types, 351; inequality of, 479, 485-87; labor's share of, 476-78; national, 196, 467-70; present value of future, 379-80 (*see also* Time preference); property's share of, 478; and law of proportionality, 437-39; a result of price, 129
See also Distribution
 Index numbers, 657-63; defined, 658-59
 Individual initiative. *See* Entrepreneur
 Individual proprietorship, 74-75
 Industrial Revolution, 55-73; in America, 71-73
 Industry, domestic system of, 51-53; factory system of, 53-54, 60-63; handicraft stage of, 43-51; household system of, 1, 40-43
 Inflation, 677-78, 683-84
 Inheritance, 473, 486, 489
 Insolvency, 88
 Insurance, 458-59; social, 488
 Interest, 329-30, 377, 379-80; defined, 329; determination of market rate of, 391-92; nature of, 397-98, 400-402; notes drawn with, 542-44; opposition to, as private income, 402; rate of, 329-35, 379; rate of, and loan market, 382; regulation of, 402-406; stability of rate, 399-400; universality of, 380
 Intermediate goods, 6, 30; production of, 31-34
 Interpolation, 148
 Invention, 34-35
 Iron, 57-58, 67, 68
 Irredeemable money systems, 687-88; history of American, 523-26, 527-33; origin of, 522-23
 Labor, a factor in production, 27, 28-29; efficiency of, 28-29, 34, 417-19; child, 70, 71; woman, 70; hours of, 70-71; conditions of, 70-71; supply of, 372, 410-11; three classes of, 407; demand for, 408, 413; bargaining power of, 419, 423, 424; immobility of, 416-17, 423, 424; defensive policies of, against unemployment, 420; changes in prosperity of, 437-39; marginal productivity of, 409-10; and law of proportionality, 439-40; and production cost, 216-17
 Laborers, classification of, 429-36; effect of Industrial Revolution on, 69-71, 440-43; varying productivity of, 417-20
 Laboring class, defined, 35
Laissez faire, 109-10, 119
See also Economic freedom
 Land, a class of capital, 29; a factor in production, 27-28; value of, based on economic rent, 362-64; value of, determined, 368-69; qualities and rents of, 366-68; urban, and diminishing returns, 365-66; urban, inequality of, 365; varying productivity of, 360-64
 Land banks, 542, 610
 Legal tender, 499-501, 531, 539; defined, 499
 Lending, 542
See also Borrowing
 Liabilities, defined, 85
 Liability, unlimited, 76; limited, 77-78, 80; double, 81
 Life insurance policy and time preference, 378 n.
 "Limping standard," 516-18, 534
 Loan contract, nature of, 381-82
 Loans, 542-45; by Bank of France, 638; by United States government, 610, 616; call, 397, 563-64; demand for, 386-91; foreign, 604-606; markets for, 392-97; payment of,

- 557-58; rate of interest on, 329, 382, 388, 396-400; supply of, 383-86; to home owners, 632-34
- Losses, 88, 242-43, 456-58
- Luxuries, 173-74; controlled by price, 128-29; and monopoly, 246-47
- Manor, 41-43
- Marginal cost, 220-25, 238-41, 355-58; curve of, related to supply curve, 256-62; defined, 221, 222; and marginal utility, 221; and marginal revenue, 236
See also Diminishing returns
- Marginal revenue, 236-41, 243, 244; defined, 237; and marginal cost, 238-39
- Marginal utility, 156, 157; and advertising, 191; balancing of, 157, 167, 205; defined, 142-43; and demand, 153-55, 167; and durable goods, 150-51; graphical representation of, 143-48; law of, 149; and marginal cost, 221; meaning of, 148-49; and price relations, 151; and supply, 151-52
- Market, town, 46-47; defined, 176; "clearing the," 272-73; open, operations in, 591-92, 626; imperfections of, 297-98; free, 273; perfectly competitive, 321-22; investment, 382; and producer, 182, 185; and seller, 182; "spoiling," 185; and supply, 265
- Marketing, evolution of, 46-50, 51-53
- Market price, 275, 297-98; and equilibrium price, 275
- Market ratio, 508-15
- Markets, classification of, 178; and competition, 183
- Mercantilism, 109, 110
- Mint ratio, 507-16; defined, 508
- Monetary systems, 501-506; present American, 535-39; British, 526-27; irredeemable, 522-39; managed, 686-88; standard, 495-520
- Money, American, in circulation, 537, 538; American stock of, 673-81; bimetallic, 507, 508, 515; classification of, 502; coin, 503-507; currency, based on, 501; defined, 21, 497; furnishes standard of value, 498; government control of, 499-500; inflation of, by gold, 675-76; history of American, 512-15, 516-19, 673-81; ideal qualities of, 504; managed, 686-88; the medium of exchange, 498; modern economy centered on, 129-30, 495; paper, 501-502, 667 n.; primitive, 495-97; quantity and interest rate, 394-96; quantity and price level, 686; quantity theory of, 668-72, 687-88; quantity and value of, 665-67, 668-81; redeemability of, 501-503, 536; representative, 501, 502-503; stable, 684-86; standard, 501-503; token, 502-503; value of, 655, 663-68, 682-84; value of, defined, 655; varieties of, 501-502; velocity of circulation, 569, 663, 667-69
- Monometallism, 505
- Monopoly, 179, 184, 188, 302; and competition, 116; and price, 245-47; and imperfect competition, 245-46; and luxuries, 246-47; and public interest, 246-47; and supply, 268
- Mutual Mortgage Insurance Fund, 634
- National banks, 573, 582, 590, 596, 622; and branch banking, 621; capital requirements, 620-21; deposit insurance for, 619-20; holding companies, 596, 599; loans on real estate, 628; notes of, 589-91; origin of system, 580-81; relation to federal reserve system, 584, 586; retirement of notes of, 581-82, 589-91
- National Credit Corporation, 609-10
- National Housing Act of 1934, 634
- National Recovery Administration, 616, 623
- Necessities, and monopoly, 4-5, 129, 173-74, 246-47
- "New Deal," 269, 626; financial philosophy of, 626
- Note broker, 571
- Notes, 83-84; federal reserve, 503, 568, 587-91, 594, 610-12, 679; government, 502, 503; national bank, 589-91; proceeds of, 544-45; promissory, 542, 545-47; secured, 545
See also Bank notes; Discount

- Oneida Community, 98
 Output, and cost, 243; and price, 243
 Overhead, 215 n., 216
See also Cost, of production
- Paper, trade, 545-47
 Paper money, 501-502, 667 n.
 Partnership, 75-77; limited, 77-78
 Pittman Act, 518-19
 Poverty, 481
 Power, 60-63
 Price, 20-21; behavior in imperfect competition, 317; causes of change in, 281; controls production, 112-14, 119, 206; and individual buyer, 177-79; consumer's influence on, 136-37, 160-62; control of, 212-13; and cost, 205-206, 210-13, 227-33; controls consumption, 112-14, 127-28; determined by bargaining, 113-14; determined by demand and supply, 164-74, 194, 204, 270-72; cutting of, factors in limitation of, 132-33, 312-17; determination and long run, 318-21; determination in short run, 312-18; distinguished from value, 22; equilibrium of, 275; fluctuation, 130-34, 189; and imperfect competition, 305-22; interrelated, 324, 337-41; law of, 276, 289-90, 299-301, 333-35; and marginal utility, 140-42, 151; and output, 233, 243; policy of, 229; and producers, 129; and profits, 464-65; and pure competition, 180-81, 270-303, 321-22; rôle of, 127-29; and single seller, 180-81, 234; setting of, 182-83; stability of, 188-89; and supply, 254-56; theory, 324, 341-43; unregulated, 116-17
 Price-fixing, 117, 133-35, 177-78
 Price level, 131-32, 663-69; factors controlling, 663
 Prices, and economic rent, 369-71; competitive, 116-17; fixed, 213; lower limit, 210-11; and irredeemable money, 680; upper limit, 211-12; reservation, 208-209; sacrifice, 209
 Production, 24-38; agricultural, 257-58, 262; and competition, 183-84; and cost, 205, 215, 216-17, 229-47; control of, 98-120; cost of, 229-47; cost of, and income, 335-36; defined, 25; equilibrium of, 295; evaluation of present system, 114-20; factors of, 24-38; financing of, 540-41; and income, 199; indirect, 31; motive of, 111-12; under imperfect competition, 233-43; under pure competition, 229-33; individuality of, 183-84; policy of, 229-47; purpose of, 129; records of, 199-201; roundabout, 31; to order, 205-206; restriction of, by workers, 437-40; unit cost of, 214
 Profit and loss statement, 93-96, 452
 Profits, 351, 452-66; actual, 454-56; criticism of system, 465-66; defined, 454; gross, 93-94; and loss, 456-58; maximum, 234-36, 239; nature of, 452-53; net, 94, 95, 453; not a price determinant, 464-65; and risk, 458-64; undivided, 87
 Promoting, 460-62
 Property, and capital, 325-26; defined, 16; division of, 17-18; distinct from wealth, 19; evidence of, 19, 82; in free persons, 18; measurement of, 20, 22; private, 110, 111, 115; its share of income, 480; valuation of, 326
 Proportionality, defined, 438; law of, and wages, 437-39; and labor, 439-40
 Proprietorship, individual, 74
- Railways, 66-67, 104-105
 Real estate, loans on, 628
 Recessions, 213
 Reconstruction Finance Corporation, 609-10, 612, 614, 615-16, 618
 Reconstruction Finance Corporation Mortgage Company, 615-16
 Rediscount, 563, 586-87, 636
 Relief, 481, 487
 Rent, 353; implicit, 453; quasi-, 371-72
See also Economic rent
 Reserve, bank, 394-95, 549; in Canada, 652; centralized, 582-84, 593-94; of federal reserve banks, 593-94; bank, influence of, 571; in France, 639, 640; function of, 560-61; of Great Britain, 642-44, 647,

- 652; lawful, 591; mobile, 594-95; ratio of, 560-66, 591
- Risk, 36-37, 397-98; profit and, 458-64; of security holders, 36-37, 84, 85
- Saving, 31-34, 36, 383-86, 473; corporate, 384-86; incentives, 32, 118-19, 383-88, 405
- Savings and Loan Insurance Corporation, 633-34
- Savings banks, 542, 574; mutual, 617, 619, 620
- Security, collateral, 545
- Seigniorage, defined, 532 n.
- Sellers, and their costs, 204-28; and demand, 176-203; individual, 305-306; liberty of, 187-88; individual, and producer, 182; policies of independent, 311-12
- Selling, 184-85, 206-10, 255-56
- Services, defined, 15; of free persons, 18, 25-26
- Sherman act (silver), 517-18
- Silver, 108; acts compelling coinage of, 517-19; free coinage of, abolished in America, 514; monetary significance of, 533-34; policy in United States, 532-33; political significance of, 534-35; price, 132; ratio of, to gold, 508-15; as standard money, 507-509; why used as money, 504-506
- Single tax movement, 374-75
- Social insurance, 385, 488
- Standard of living, 443-46
- Stock. *See* Capital stock
- Stock exchange, 299-301
- Stock of goods, not supply, 252-53
- Strikes, and wages, 414
- Substitutes, demand for, 307; and demand, 308-10; nature of, 308; significance of, 306-308
- Substitution, 167-68, 188, 246; and demand, 172-73
- Supply, analogous to demand, 252; and demand, 204-205, 265; attempts to control, 268-69; changes in, 284-87, 290-93; concepts of, 255, 279-81; definition of, 280; controlled by cost, 256-62; curves, 254-55, 260-63; elasticity of, 260-61, 287-89; importance of, 265; indexes of potential, 267-68; individual, defined, 250; individual, and price, 250-51; laws of, stated, 261, 263, 275; long run and short run, 261-62; measurement of potential, 265, 266-68; nature of, 262; and price, 254; and stock of goods, 252-53; and pure competition, 254-55; total, defined, 252
- Surplus, 87; in property account, 87-88; permits cooperation, 104; producers, 358-60, 363, 372-73
- Tariff, protective, 302
- Taxation, of economic rent, 374-75
- Textiles, 58-64, 68; in America, 72
- Thomas amendment, 530
- Time preference, 377-79; and borrowing for consumption, 389-90; and capital production, 379; defined, 377; determines consumption borrowing, 389-90; and life insurance, 378 n.; rate of, 379
See also Income, present value of future
- Town, medieval, 45-49, 50-51
- Trade, beginnings of, 45-47; balance of, 108, 108 n.; free, 47-48; at medieval fairs in England, 68, 109; relation of, to quantity of currency, 671-72; value of money varies with, 665, 668, 674-75
See also Exports; Free Trade, Imports
- Trade marks, 184
- Trade paper, types of, 545-46
- Trade unions, and collective bargaining, 419; wages, rates, 419-20; restrictive tendencies of, 419-20; and wages, 117, 416
- Transfer, defined, 20-21
- Transportation, 64-67, 105, 338
- Treasury, United States, 579-81, 596-97
- Trust companies, 542, 573-74, 581
- Underwriting, 460-61
- Unearned increment, 373-74, 375
See also Economic rent
- Unemployment, 684; and minimum wages, 450-51; and trade unions, 419-20
See also Employment
- Unions. *See* Trade unions

- Usury, 403-404; laws of, 403, 406
- Utility, balancing of, 157, 160, 167; creation of, 24-26; defined, 9-10, 16; marginal, and demand, 153-55, 167; forms of, 26; marginal, 137-43; measurement of, 156; nature of, 139; subjective side of, 137-39; total, 150
See also Marginal utility
- Utilization, margins of, 353-55
- Value, book, 89, 343-44, 345; common unit of measure, 21, 22; defined, 21; distinct from price, 22; and marginal utility, 151-52; market, 89, 344-46; of money, and measurement of, 654-57, 657-63; par, 89; present, of future income, 324-25 (*see also* Time preference); rate base, 344-45; the resultant of judgment, 343-46; tax, 345; theory of, 341-46
- Wages, children's, 448; defined, 407; forces determining, 407-11; and immigration, 480; of independent workers, 426-27; and labor strikes, 414; law of, 411-13, 422, 426, 448-51; and law of proportionality, 438-40; of management, 210, 211, 398, 453-54; and marginal productivity, 409-10, 414-17, 425-26, 450; minimum, and legislation concerning, 448-51; piece rate, 421-22; policy of high, 445, 446-48; the price of labor, 112, 133, 407; rates of, 133; real, 436-40, 683, 684; regulation of, 107, 417, 446-48; sphere of bargaining in, 413-17; and standard of living, theory of, 443-46; systems of payment, 420-22; task and bonus systems, 421-22; on time base, 420; and trade unions, 117, 416; variations in, 417-19, 422-25, 429-34; woman's, 418, 423, 448
- Wages of management, 210, 211, 398, 453-54
- Walker, Francis A., quoted, 9
- Wants, 3-4; divided into necessities and luxuries, 4-6; utility the satisfaction of, 9-10; without limit, 6-7
- War debts, 605-606
- War Finance Corporation, 615
- Wealth, costs of, 15; defined, 10-14; distinct from property, 16; income of, 15-16; measurement of, 20; net income of, 15-16; ownership of, 16; production of, 25-27
- Woollen industry, 58-60, 61